

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:40:33 ; Search time 56.6732 Seconds
(without alignments)
1736.455 Million cell updates/sec

Title: US-09-941-947a-6

Perfect score: 3216
Sequence: 1 MKLTDPYPLKNTHTPADIR.....LSLVGSDSKGLATIEQFCA 620

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3216	100.0	620	23	ABG61581
2	3216	100.0	620	23	AAE22301
3	3216	100.0	620	23	AAU80325
4	2069.5	64.4	627	22	AAU00012
5	2069.5	64.4	627	23	AAO21846
6	2063.5	64.2	627	22	AAU00014
7	2061.5	64.1	627	22	AAU00013
8	2014.5	62.6	620	22	AAU8450
9	2000.5	62.2	620	20	AA52832

10	2000.5	62.2	620	20	AAV08880
11	2000.5	62.2	620	21	AAE20544
12	2000.5	62.2	620	21	AA51613
13	2000.5	62.2	620	22	AAU34475
14	2000.5	62.2	620	23	AAO21860
15	1937.5	60.2	625	23	AAO21862
16	1776	55.2	637	23	AAO21861
17	1774	55.2	637	24	ABP7532
18	1620.5	50.4	628	23	AAO21847
19	1532.5	47.7	641	23	AAO21845
20	1504.5	46.8	648	20	AAV52834
21	1491.5	46.4	648	23	AAO21856
22	1485.5	46.2	720	21	AAV97422
23	1442	44.8	640	23	AAO21858
24	1439	44.7	636	23	AAO21857
25	1420.5	44.2	708	21	AAV97415
26	1418.5	44.1	724	20	AAV14145
27	1418.5	44.1	727	20	AAV14146
28	1411.5	43.9	637	23	AAO21855
29	1411.5	43.9	659	23	AAO48246
30	1411.5	43.9	717	21	AAO42534
31	1411.5	43.9	717	23	AAO48245
32	1411.5	43.9	824	23	AAO48247
33	1410.5	43.9	641	23	AAO21854
34	1410.5	43.9	721	21	AAV97414
35	1407.5	43.8	725	21	AAV15157
36	1406.5	43.7	637	20	AAV52833
37	1404.5	43.7	717	21	AAV51612
38	1404.5	43.7	717	21	AAV51667
39	1402.5	43.6	719	23	AAO21851
40	1402	43.6	735	23	AAO21849
41	1394.5	43.4	633	21	AAO21848
42	1394.5	43.4	633	23	AAO21843
43	1394.5	43.3	633	23	AAO21848
44	1392	43.3	721	20	AAV14144
45	1389.5	43.2	632	21	AAO42535

ALIGNMENTS

RESULT 1
ID ABG61581 standard; Protein: 620 AA.
XX
AC ABG61581;
XX
DT 27-AUG-2002 (first entry)
XX
DB High growth mechanotrophic bacterial strain polypeptide #21.
XX
XX High growth mechanotrophic bacterial strain; C1 carbon substrate; enzyme;
XX methan; methanol; Embden-Meyerhof carbon flux pathway; 16S RNA;
XX pyrophosphate dependent phosphofructokinase; nitrogen-containing compound;
XX ammonia; nitrate; nitrite; nitrogen; pigment; oxygen; landfill;
XX methane-containing environment; waste water treatment system; isoprenoid;
XX nitrous oxide; terpenoid; animal feed; carotenoid; exopolysaccharide.
XX
XX Methylobionas 16a.
XX
XX MO200220728-A2.
XX
XX
XX PD 14-MAR-2002.
XX
XX 28-AUG-2001; 2001WO-US26827.
XX PF
XX 01-SEP-2000; 2000US-229858P.
XX PR
XX (DUPO) DU PONT DE NEMOURS & CO E. I.
XX
XX Koffas M, Odom JM, Schenkele A;
XX
XX WPI; 2002-452200/48.
XX
XX

E. coli DXS protei
Escherichia coli D
E. coli DXS prote
E. coli cellular p
Isoprenoid related
Isoprenoid related
Isoprenoid related
N. gonorrhoeae am
Isoprenoid related
Isoprenoid related
Rhodobacter sphaer
Isoprenoid related
Rice 1-deoxy-D-xyl
Isoprenoid related
Isoprenoid related
Soybean 1-deoxy-D-
M. piperita 1-deox
M. piperita 1-deox
Isoprenoid related
Truncated deoxyxyl
Arabidopsis thalia
Arabidopsis thalia
Thioredoxin/deoxyx
Isoprenoid related
Soybean 1-deoxy-D-
Tagetes erecta put
Rhodobacter sphaer
A. thaliana DXS p
A. thaliana DXS p
Isoprenoid related
Isoprenoid related
Synchocystis sp.
Bacillus subtilis
Isoprenoid related
M. piperita 1-deox
Arabidopsis thalia

DR N-PSDB; ABR3260.

XX New high growth methanotrophic bacterial strain, useful for producing
PT single cell proteins, grows on a C1 carbon substrate, and comprises a
PT functional gene encoding in Embden-Meyerhof carbon pathway
XX
PS Claim 11; Page 138-140; 157pp; English.

XX The invention relates to a high growth methanotrophic bacterial strain,
CC which grows on a C1 carbon substrate e.g. methane and methanol, and
CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a
CC gene coding a pyrophosphate dependent phosphofructokinase enzyme or a 16S
CC RNA. The bacterial strain is useful for the production of single cell
CC protein and for the biotransformation of a nitrogen-containing compound,
CC e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the
CC production of a feed product comprising a protein, carbohydrates and a
CC pigment and for reducing oxygen demand, for removing nitrates and
CC nitrates in methane-containing environments such as landfills, waste
CC water treatment systems or anywhere that methane, oxygen and nitrates are
CC present. The bacterial strain of the invention can be used as a
CC denitrifying agent for the conversion of nitrate or nitrite to nitrous
CC oxide with methane or methanol as a carbon source. It is also used in the
CC production of biomass including proteins, carbohydrates and a wide
CC variety of pigments (particularly for isoprenoid pigments for the
CC purpose of generating animal feeds), in production of terpenoid and
CC carotenoid compounds, useful as pigments and as monomers in polymeric
CC materials and in production of exopolysaccharides at high levels.
CC Sequences ABR6151-ABR61590 represent high growth methanotrophic
CC bacterial strain proteins of the invention.

XX Sequence 620 AA;

XX Query Match 100.0%; Score 3216; DB 23; Length 620;

XX Best Local Similarity 100.0%; Pred. No. 1.2e-285; Indels 0; Gaps 0;

XX Matches 620; Conservative 0; Mismatches 0;

XX 1 MKLTTDYPPLKNTHTPADIRALSKDQLQADVRGYLTHTVTSISGHPAAGGTVELTV 60

XX 1 MKLTTDYPPLKNTHTPADIRALSKDQLQADVRGYLTHTVTSISGHPAAGGTVELTV 60

XX 61 ALHYVPTPVQDLVWDVGHQAYPKKILITGRKEMPTTRTGYSAPFAPABESYDAFGVG 120

XX 61 ALHYVPTPVQDLVWDVGHQAYPKKILITGRKEMPTTRTGYSAPFAPABESYDAFGVG 120

XX 61 ALHYVPTPVQDLVWDVGHQAYPKKILITGRKEMPTTRTGYSAPFAPABESYDAFGVG 120

XX 121 HSTSTISAAAGMALASOLRGEDKKMAVLIIGDSTTGGAAYEAMHAGDVANLILVINDN 180

XX 121 HSTSTISAAAGMALASOLRGEDKKMAVLIIGDSTTGGAAYEAMHAGDVANLILVINDN 180

XX 121 HSTSTISAAAGMALASOLRGEDKKMAVLIIGDSTTGGAAYEAMHAGDVANLILVINDN 180

XX 181 DMSISPPVGANNVLTLYLSKSYSSVRESKKAALAMPVWEIARTEEHVYKATVPGT 240

XX 181 DMSISPPVGANNVLTLYLSKSYSSVRESKKAALAMPVWEIARTEEHVYKATVPGT 240

XX 181 DMSISPPVGANNVLTLYLSKSYSSVRESKKAALAMPVWEIARTEEHVYKATVPGT 240

XX 241 LPEELGFRVYFPPIDGHVEMVSTLENKDETSVPVPLAVTKKKGAPAKEDPLAHGV 300

XX 241 LPEELGFRVYFPPIDGHVEMVSTLENKDETSVPVPLAVTKKKGAPAKEDPLAHGV 300

XX 241 LPEELGFRVYFPPIDGHVEMVSTLENKDETSVPVPLAVTKKKGAPAKEDPLAHGV 300

XX 301 PAPPTKDFLKPAAAPSPHTYTFEYGRMLCDMAADDERLIGITPAMEGSGVFEFSQFP 360

XX 301 PAPPTKDFLKPAAAPSPHTYTFEYGRMLCDMAADDERLIGITPAMEGSGVFEFSQFP 360

XX 301 PAPPTKDFLKPAAAPSPHTYTFEYGRMLCDMAADDERLIGITPAMEGSGVFEFSQFP 360

XX 361 NRYFDVAIAEGHANTVLAAGACOCAGKPVVALYSTTLOKNGOULHDVALQMLDMLFALDR 420

XX 361 NRYFDVAIAEGHANTVLAAGACOCAGKPVVALYSTTLOKNGOULHDVALQMLDMLFALDR 420

XX 361 NRYFDVAIAEGHANTVLAAGACOCAGKPVVALYSTTLOKNGOULHDVALQMLDMLFALDR 420

XX 421 AGVGPDPPTAGAFDYSYMRCPNMLIMAPADENECROMLTGPOHHGPAVYEPKGG 480

DB 541 ALVLELARTHDFVTEVENVIAGAGSAINTEFLQAKVLMPEVCNIGLDFRVEQGSREEL 600

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

DB 601 LSLVGLDSKGLIATIEQPCA 620

QY 61 ALHYVNTPVQOLVWDVGHQAYPHKILTGKREMPRTITLGGVSAFPARDSEYDAFGV 120
 DB 61 ALHYVNTPVQOLVWDVGHQAYPHKILTGKREMPRTITLGGVSAFPARDSEYDAFGV 120
 QY 121 HSTSTISALGMAIASQLRGEDKQVVAIGDGSITGGMAVEANRAGVANNLVTINDN 180
 DB 121 HSTSTISALGMAIASQLRGEDKQVVAIGDGSITGGMAVEANRAGVANNLVTINDN 180
 QY 181 DMSISPPVGMANNLTLYKLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPCT 240
 DB 181 DMSISPPVGMANNLTLYKLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPCT 240
 QY 241 LPEELGFNFEGPIGDHVEMLVSTLENLKDLPVFLHVYTKKGYAPAEKOPLAYHGV 300
 DB 241 LPEELGFNFEGPIGDHVEMLVSTLENLKDLPVFLHVYTKKGYAPAEKOPLAYHGV 300
 QY 301 PAFDPTKDFLEKAPSPHPTYTEVFGKMLCDMAADDERLLGITPAMRBSGLVBSQKFP 360
 DB 301 PAFDPTKDFLEKAPSPHPTYTEVFGKMLCDMAADDERLLGITPAMRBSGLVBSQKFP 360
 QY 361 NRYFDVAIAEQHAYVTLAAGACQAKPVVAISTFLQRYDQILHDVALQNLDMFLALDR 420
 DB 361 NRYFDVAIAEQHAYVTLAAGACQAKPVVAISTFLQRYDQILHDVALQNLDMFLALDR 420
 QY 421 AGLVGPDGFTTHAGAFDYSYMRCTPNNLIMAPADENECRQMLTTGFOHGPASVRYPRGK 480
 DB 421 AGLVGPDGFTTHAGAFDYSYMRCTPNNLIMAPADENECRQMLTTGFOHGPASVRYPRGK 480
 QY 481 PGAAIDPITLALBEGKAEVREHGSRIAILAMGSMTVPAVEAGKOLGATVNNRFXKPEDQ 540
 DB 481 PGAAIDPITLALBEGKAEVREHGSRIAILAMGSMTVPAVEAGKOLGATVNNRFXKPEDQ 540
 QY 541 ALVLELARTHDVFTVEENVVLAGAGSAINTEFLQAKVLMPCVNIQLPDRFVEQSSREEL 600
 DB 541 ALVLELARTHDVFTVEENVVLAGAGSAINTEFLQAKVLMPCVNIQLPDRFVEQSSREEL 600
 QY 601 LSLVGLDSKGLIATIEQFCA 620
 DB 601 LSLVGLDSKGLIATIEQFCA 620

RESULT 3
 AAU80325
 ID AAU80325 standard; Protein; 620 AA.
 XX AC AAU80325;
 XX DT 15-JUN-2002 (first entry)
 XX DE Methylomonas 16a ORF1 dks (1-deoxyxylulose-5-phosphate synthase) protein.
 XX KM Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;
 XX KW kerateneid; pigment; flavour; fragrance; open reading frame 1; ORF1.
 XX dks; 1-deoxyxylulose-5-phosphate synthase enzyme.
 XX OS Methylomonas sp.
 XX PN WO200220733-A2.
 XX PD 14-MAR-2002.
 XX PF 29-AUG-2001; 2001WO-US26852.
 XX PR 01-SEP-2000; 2000US-229907P.
 XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX PI Cheng Q, Koffas M, Norton KC, Odom JM, Picataggio SK, Rouviere PB;
 XX PI Schenzzle A, Tomb J;
 XX WPI; 2002-383051/41.
 XX DR N-PSDB; AER50081.
 DR

XX Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,
 PT isolated from Methylomonas 16a, useful for the production of isoprenoid
 PT compounds -
 XX
 XX PS
 XX Claim 4; Page 66-68; 84pp; English.
 CC The present invention relates to a new nucleic acid molecule encoding
 CC an isoprenoid biosynthetic enzyme isolated from Methylomonas 16a.
 CC The invention is useful for obtaining a nucleic acid molecule
 CC encoding an isoprenoid compound biosynthetic enzyme, and for the
 CC microbial production of isoprenoid compounds. The molecules of the
 CC invention are also useful for regulating isoprenoid biosynthesis in an
 CC organism and for producing recombinant organisms for producing various
 CC isoprenoid compounds. The nucleic acid is also useful for feed additive,
 CC for the production of kerateneids and their derivatives, isoprenoid
 CC intermediates, and as pure products useful as pigments, flavours and
 CC fragrances. The present amino acid sequence represents the Methylomonas
 CC 16a open reading frame 1 (ORF1) dks (1-deoxyxylulose-5-phosphate synthase
 CC enzyme) protein of the invention, as described above.

Sequence 620 AA;

Query Match 100.0%; Score 3216; DB 23; Length 620;
 Best Local Similarity 100.0%; Pred. No. 1.2e-285;
 Matches 620; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLITDYPILANKITPADIPLALSKDQLOQLADYVRGTLTHVTSISGGHFAAGLGVETLV 60
 DB 1 MKLITDYPILANKITPADIPLALSKDQLOQLADYVRGTLTHVTSISGGHFAAGLGVETLV 60
 QY 61 ALHYVNTPVQOLVWDVGHQAYPHKILTGKREMPRTITLGGVSAFPARDSEYDAFGV 120
 DB 61 ALHYVNTPVQOLVWDVGHQAYPHKILTGKREMPRTITLGGVSAFPARDSEYDAFGV 120
 QY 121 HSTSTISALGMAIASQLRGEDKQVVAIGDGSITGGMAVEANRAGVANNLVTINDN 180
 DB 121 HSTSTISALGMAIASQLRGEDKQVVAIGDGSITGGMAVEANRAGVANNLVTINDN 180
 QY 181 DMSISPPVGMANNLTLYKLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPCT 240
 DB 181 DMSISPPVGMANNLTLYKLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPCT 240
 QY 241 LPEELGFNFEGPIGDHVEMLVSTLENLKDLPVFLHVYTKKGYAPAEKOPLAYHGV 300
 DB 241 LPEELGFNFEGPIGDHVEMLVSTLENLKDLPVFLHVYTKKGYAPAEKOPLAYHGV 300
 QY 301 PAFDPTKDFLEKAPSPHPTYTEVFGKMLCDMAADDERLLGITPAMRBSGLVBSQKFP 360
 DB 301 PAFDPTKDFLEKAPSPHPTYTEVFGKMLCDMAADDERLLGITPAMRBSGLVBSQKFP 360
 QY 361 NRYFDVAIAEQHAYVTLAAGACQAKPVVAISTFLQRYDQILHDVALQNLDMFLALDR 420
 DB 361 NRYFDVAIAEQHAYVTLAAGACQAKPVVAISTFLQRYDQILHDVALQNLDMFLALDR 420
 QY 421 AGLVGPDGFTTHAGAFDYSYMRCTPNNLIMAPADENECRQMLTTGFOHGPASVRYPRGK 480
 DB 421 AGLVGPDGFTTHAGAFDYSYMRCTPNNLIMAPADENECRQMLTTGFOHGPASVRYPRGK 480
 QY 481 PGAAIDPITLALBEGKAEVREHGSRIAILAMGSMTVPAVEAGKOLGATVNNRFXKPEDQ 540
 DB 481 PGAAIDPITLALBEGKAEVREHGSRIAILAMGSMTVPAVEAGKOLGATVNNRFXKPEDQ 540
 QY 541 ALVLELARTHDVFTVEENVVLAGAGSAINTEFLQAKVLMPCVNIQLPDRFVEQSSREEL 600
 DB 541 ALVLELARTHDVFTVEENVVLAGAGSAINTEFLQAKVLMPCVNIQLPDRFVEQSSREEL 600
 QY 601 LSLVGLDSKGLIATIEQFCA 620
 DB 601 LSLVGLDSKGLIATIEQFCA 620

RESULT 4

AA00012
 ID AA00012 standard; Protein; 627 AA.
 XX
 AC AA00012;
 DT 08-MAY-2001 (first entry)
 DE Pseudomonas aeruginosa Deoxyxylulose-5-phosphate synthase, dxs.
 XX
 KW Deoxyxylulose-5-phosphate synthase; dxs; agonist; antagonist;
 KW diagnostic; eukaryotic extracellular matrix protein; antibody; wound;
 KW in-dwelling device; Helicobacter pylori infection; gastric ulcer;
 KW gastritis; gastrointestinal carcinoma.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN WO200114409-A1.
 PD 01-MAR-2001.
 PF 16-AUG-2000; 2000MO-US22332.
 PR 19-AUG-1999; 99US-0377279.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Huang J, Jiang X, McDevitt D;
 XX
 DR WPI: 2001-211303/21.
 DR N-PSDB; AAS00017.
 XX
 PT New deoxyxylulose-5-phosphate (dxs) polypeptides of Pseudomonas
 PT aeruginosa useful for diagnosis of dxs expression or activity related
 PT disease -
 XX
 PS Claim 1; Page 4; 41pp; English.
 XX
 CC The sequence represents Pseudomonas aeruginosa deoxyxylulose-5-phosphate
 CC synthase, dxs. (Ant)agonists of dxs are useful for treating an individual
 CC in need of enhanced or inhibited activity or expression of or
 CC immunological response to dxs. Dxs and the nucleic acid encoding it are
 CC useful as diagnostic reagents, especially in relation to diseases related
 CC to expression or activity of dxs. Fragments of the nucleic acid are
 CC useful as probes or primers and to synthesize full length dxs
 CC polynucleotides. The nucleic acid and protein are useful for diagnosing
 CC bacterial infections, especially by Pseudomonas aeruginosa, staging a
 CC disease or response of an infectious organism to drugs, and are also used
 CC in a structure based design of an (ant)agonist to dxs. The nucleic acid,
 CC protein and antibodies raised against it are useful for screening for
 CC (ant)agonists of dxs or the nucleic acid and are useful to interfere with
 CC the initial physical interaction between a pathogen and a mammalian host
 CC responsible for sequelae of infection. These molecules are useful in
 CC preventing adhesion of gram positive and/or gram negative bacteria to
 CC eukaryotic extracellular matrix proteins or in-dwelling devices or in
 CC wounds; to block bacterial adhesion; to block the normal progression of
 CC pathogenesis in infections initiated other than by the implantation of
 CC in-dwelling devices or by other surgical techniques. Dxs is also useful
 CC for identifying membrane bound or soluble receptors. The (ant)agonists
 CC are useful for treating Helicobacter pylori infection, and also to
 CC prevent, inhibit or treat gastric ulcers, gastritis and gastrointestinal
 CC carcinoma.
 CC
 XX
 SQ Sequence 627 AA;

Query Match 64.4%; Score 2069.5; DB 22; Length 627;
 Best Local Similarity 63.4%; Pred. No. 1.7e-180;
 Matches 388; Conservative 100; Mismatches 117; Indels 7; Gaps 2;

QY 8 PLKRIHTPADIRALSKDLOQLADEVRGYLTHYISIGSGHRAAGLGVETLVATRYEN 67
 DB 16 PLDRASSPAERLRUGADLFTLADLRQYLLYTVGQGTGHRGAGLGVETLVATRYEN 75

QY 68 TPVDQLVWDVGHQAVPEKILTGKREMEPTITLGGVSAFPARDSEVDAGVGHSTIS 127
 DB 76 TPDDRVLVWDVGHQAVPEKILTERRELMGTLQKNGLAFFPRARSEYDTFGVGHSTIS 135
 QY 128 AALGKALASOLRGDKWVALIGDSITCGMAVEAMONHAGVYNNLVLTLNDNMSTSP 187
 DB 136 AALGKALASOLRGDKWVALIGDSITCGMAVEAMONHAGVYNNLVLTLNDNMSTSP 195
 QY 188 VQAMNNVLTVALSKFYSVVEESKKAAMPVWEALAKTEEHVKMIVGTLPEELGF 247
 DB 196 VGSLSNYLAKLISRTYSMEKSKVLSRLPGAMEIARTEEYAKGLVPGTLPEELGW 255
 QY 248 NYRPIGCHDVEMVSTLENKDLTGPEVLTWTKKGGVAPAKKDLAVGVAFPDPTX 307
 DB 256 NYRPIGCHDVEMVSTLENKDLTGPEVLTWTKKGGVAPAKKDLAVGVAFPDPTX 310
 QY 308 DFLPYAAP--SPHPTYTEVFGRMLCDMAADERLLGITPARREGSGVFEQKPNRYFD 365
 DB 311 LEAPGSAFKTKGGRKYSVFGOMLCDMAADARLLGITPARREGSGDVAFSERPERYFD 370
 QY 366 VALABQHAVTTLAQAQACQAKPVAIYSTTLQSGIDQLHDVALQNDMLPALDRAGIV 425
 DB 371 VALABQHAVTTLAQAQACQAKPVAIYSTTLQSGIDQLHDVALQNDMLPALDRAGIV 430
 QY 426 PDGPTHAGAPFYSYMRCLPMTLIMAPDENECRQMLTTGQHGHPASVYPRGKPGAAI 485
 DB 431 EDGPTHAGAPFYSYMRCLPMTLIMAPDENECRQMLTTGQHGHPASVYPRGKPGAAI 490
 QY 486 DFTLTALRIGKAEVRRHGSRIALIAMGSMVTPAVEACKQGLGAVVNNRFPVFPQALYLE 545
 DB 491 DPDLQPEVIGKGVRRRGRVALLVFGVQLAEAKRVASLDATVAVDMRFVFPDLDEALVRE 550
 QY 546 LARHNDPVVVEENVVIGAGSAINTELQAKVAMPVPCNIGLPDRFVQSGREELISLV 605
 DB 551 LAGSHLVTTEENAVMGAGSAGVETLASGLVLPQLQELPDVYVHAAPSEMLAECG 610
 QY 606 LDSKGIATTEQ 617
 DB 611 IDAAGIRKAVQ 622

RESULT 5
 AA021846
 ID AA021846 standard; Protein; 627 AA.
 XX
 AC AA021846;
 DT 13-SEP-2002 (first entry)
 DE Isoprenoid related protein sequence SEQ ID No 15.
 XX
 KW Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;
 KW decaprenyl diphosphate synthase.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN WO200226933-A2.
 PD 04-APR-2002.
 PR 28-SEP-2001; 2001MO-US30328.
 XX
 PR 29-SEP-2000; 2000US-236580P.
 XX
 PA (CRGI) CARGILL INC.
 PA
 PI Gokarn R, Jessen H, Zidwick MJ;
 XX
 DR WPI: 2002-416480/44.
 XX
 PT Substantially pure polypeptides having e.g.,
 PT 1-deoxyxylulose-5-phosphate synthase activity, useful for the
 PT production of isoprenoids, especially CoQ(10)

XX Disclosure: Fig 6: 246pp: English.

PS The invention relates to methods and materials for the production of
 CC isoprenoids. More particularly the invention provides isolated nucleic
 CC acids, substantially pure polypeptides, host cells, and methods for
 CC producing various isoprenoid compounds. The polypeptides are useful for
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (Dxs)
 CC activity or decaprenyl diphosphate synthase (Dds) activity, is useful for
 CC increasing production of CoQ(10) in a cell having endogenous Dds
 CC activity. This sequence represents a protein relating to the isoprenoid
 CC production of the invention.

XX Sequence 627 AA:

Query Match 64.4%; Score 2069.5; DB 23; Length 627;
 Best Local Similarity 63.4%; Pred. No. 1.7e-180;
 Matches 388; Conservative 100; Mismatches 117; Indels 7; Gaps 2;

8 PLKNIHTPADIRALSKDQQLADEVRGYLTHTVTSISGHPAAGTGYELTVALHYVN 67
 16 PLIDRASSPAELRRKADLETADLRQLYTVQGTGHPAGAGVVELTIALHYVD 75
 68 TPVDQLVMDVGHQAYPHKILITGRKEMPTIRTLGVSAPFAPDESEYDAFGVSHSSTIS 127
 76 TPDDRLVMDVGHQAYPHKILITGRKEMPTIRTLGVSAPFAPDESEYDAFGVSHSSTIS 135
 128 PALGMAIASQLGDEKQVAIIIGDSITGMAVEANNAAGVNAVLVINDNKSISPP 187
 136 PALGMAIASQLGDEKQVAIIIGDSITGMAVEANNAAGVNAVLVINDNKSISPP 195
 188 VGMANNVYITKVLSSKFSYSSRESEKALAKMPSVEIARKTEHYKMTVPPTLFEELGF 247
 196 VGLSNVYITKVLSSKFSYSSRESEKALAKMPSVEIARKTEHYKMTVPPTLFEELGF 255
 248 NTPGPIHGDVEMVSTLENKDLTGVPVFLHYVTKKGVPADEKPLAYHGVAPDPTK 307
 256 NTPGPIHGDVEMVSTLENKDLTGVPVFLHYVTKKGVPADEKPLAYHGVAPDPTK 310
 308 DPLKPAAP--SPHPYTYEYGRWLCDMAQDSERLIGITPMMEGSGVLEESQKPEPNRYD 365
 311 LEAPGSAPEKKTGPKYSSVFGQNLCDMAQDSERLIGITPMMEGSGVLEESQKPEPNRYD 370
 366 VALAEGHAYTTLAAGACQAGAKPEVVAIYSTFLQRYDQDLIHVDALQNLDTLFDALDAGLVG 425
 371 VALAEGHAYTTLAAGACQAGAKPEVVAIYSTFLQRYDQDLIHVDALQNLDTLFDALDAGLVG 430
 426 PDGPIHAGFDVSYNKCTIPNMLIMAPADENECQMLTTGFQHHGASAVRYPRGKGPAAI 485
 431 EDGPIHAGFDVSYNKCTIPNMLIMAPADENECQMLTTGFQHHGASAVRYPRGKGPAAI 490
 486 DFTLTALEIGKAEVHHHSRIATILAMGSVTEPAVAGKQLGATVNMVFVPPFDQALVLE 545
 491 DDDIAPVEIGKAVRRRGRVALLVFGVQLAEMMVASLDATVDMFVFPDLDEALVRE 550
 546 IARTIDVAVTYEENVYIAGAGSAINITFLQAGKVLMPVCNIGLPRFVYQSGSREBLISLVG 605
 551 IAGSHEILVYTIENAVMGAGSAGVSEFLASBEFLVPLQLGLPDYVYHAKPSEMLACBG 610
 606 LDKKILATIRG 617
 611 LDAGIERAVRQ 622

RESULT 6
 AAU00014 standard: Protein; 627 AA.

XX AC AAU00014;
 XX 08-MAY-2001 (first entry)
 DT
 XX

DB Pseudomonas aeruginosa dxs ts-80 mutant.

XX Deoxyxylulose-5-phosphate synthase; dxs; agonist; antagonist;
 XX diagnostic; eukaryotic extracellular matrix protein; antibody; wound;
 XX in-dwelling device; Helicobacter pylori infection; gastric ulcer;
 XX gastritis; gastrointestinal carcinoma; mutant; mutein; ts-32.

OS Pseudomonas aeruginosa.

XX Key Location/Qualifiers

XX Misc-difference 380 /note="Wild-type Thr replaced by Ile"

XX MO20011409-A1.

XX 01-MAR-2001.

XX 16-AUG-2000; 2000WD-US22332.

XX 19-AUG-1999; 99US-0377279.

XX (SMK) SMITHKLINE BEECHAM CORP.

XX (SMK) SMITHKLINE BEECHAM PLC.

XX Huang J, Jiang X, McDevitt D;

XX MPI; 2001-211303/21.

XX N-PSDB; AAS00019.

XX New deoxyxylulose-5-phosphate (dxs) polypeptides of Pseudomonas

XX aeruginosa useful for diagnosis of dxs expression or activity related

XX disease -

XX Example 2; Page -; 41pp; English.

XX The sequence represents Pseudomonas aeruginosa deoxyxylulose-5-phosphate
 CC synthase, dxs, temperature sensitive mutant, ts-80. (Ant)agonists of dxs
 CC are useful for treating an individual in need of enhanced or inhibited
 CC activity or expression of or immunological response to dxs. Dxs and the
 CC nucleic acid encoding it are useful as diagnostic reagents, especially in
 CC relation to diseases related to expression or activity of dxs. Fragments
 CC of the nucleic acid are useful as probes or primers and to synthesize
 CC full length dxs polynucleotides. The nucleic acid and protein are useful
 CC for diagnosing bacterial infections, especially by Pseudomonas
 CC aeruginosa, staging a disease or response of an infectious organism to
 CC drugs, and are also used in a structure based design of an (ant)agonist
 CC to dxs. The nucleic acid, protein and antibodies raised against it are
 CC useful for screening for (ant)agonists of dxs or the nucleic acid and are
 CC useful to interfere with the initial physical interaction between a
 CC pathogen and a mammalian host responsible for sequelae of infection.
 CC These molecules are useful in preventing adhesion of gram positive and/or
 CC gram negative bacteria to eukaryotic extracellular matrix proteins or
 CC in-dwelling devices or in wounds; to block bacterial adhesion; to block
 CC the normal progression of pathogenesis in infections initiated other than
 CC by the implantation of in-dwelling devices or by other surgical
 CC techniques. Dxs is also useful for identifying membrane bound or soluble
 CC receptors. The (ant)agonists are useful for treating Helicobacter pylori
 CC infection, and also to prevent, inhibit or treat gastric ulcers,
 CC gastritis and gastrointestinal carcinoma.
 CC The present sequence is not shown in the specification but is derived
 CC from the dxs sequence given in AAU00012.

XX Sequence 627 AA;

Query Match 64.2%; Score 2063.5; DB 22; Length 627;
 Best Local Similarity 63.2%; Pred. No. 6.1e-180;
 Matches 387; Conservative 100; Mismatches 118; Indels 7; Gaps 2;

XX 8 PLKNIHTPADIRALSKDQQLADEVRGYLTHTVTSISGHPAAGTGYELTVALHYVN 67
 DB 16 PLIDRASSPAELRRKADLETADLRQLYTVQGTGHPAGAGVVELTIALHYVD 75
 68 TPVDQLVMDVGHQAYPHKILITGRKEMPTIRTLGVSAPFAPDESEYDAFGVSHSSTIS 127

```

Db 76 TDDRLVMDVGHQAVPHKILTERBELMGLRQKNGLAAPRRAESEYDFRGHSTSTIS 135
Qy 128 AALGMAIASQLGEGDKKKVAIIGDSITGMAAYEAMNAGDVANLIVLINDMSISPP 187
Db 136 AALGMAIAARLOQEKERSVAIVGDGALTAGMAFEALNHASEVDADMEVLINDMSISHN 195
Qy 188 VGAMNNYLTKVLSKREYSSVREESKKALAMPVWEIARKEEYKGMIVPGTLFEEIGF 247
Db 196 VGGLSNTYLAKLISRRYSRMSRSGSKVLSRLPGAMBIARTETEVAKKGLVPTLLEELGM 255
Qy 248 NYFGPIDGHVEMLVSTENLKDITGVPVLEVVYTKKGAVAPAEKDPPLAYHGVPAFDPK 307
Db 256 NYIGPIDGHDLPTLVATLNNRDMKGPQFLHVVTKKGKGFAPALDPIGYHAI-----TK 310
Qy 308 DFLPKAAP--SPHPTTVYVGRMLCDMAADERLIGITPAMEGSGLVESGKZPNRYPD 365
Db 311 LEAPGSAFKKGTGPKTSVFGQMLCDMAADARLLGITPAMEGSDLVAFSERYPERYFD 370
Qy 366 VALAQHAAVTLAAGACQAGAKPVAIYSTFLQRYDOLIHVVALQNLDMPLALDRAGLVG 425
Db 371 VALAQHAAVTLAAGACQAGAKPVAIYSTFLQRAYDOLIHVAVQHLDLVFAIDRAGLVG 430
Qy 426 PGGPTTAGAFDYSYKACIPNMLIMAPADNECQMLTTFQFHGPASTRYPRGKPGALAI 485
Db 431 EGGPTTAGAFDYSYKACIPNMLIMAPADNECQMLTTFQFHGPASTRYPRGKPGALAI 490
Qy 486 DFLTALAEIGKAEVHHSGRIALLAMGWSVTPAVEAGKQATVNMKPFVFPDQALVLE 545
Db 491 DEDLOVDELGKGVNRGRGRVALLVFGVGLAEMAMVASLIDATVDMFVFDLBAVRE 550
Qy 546 LARTEDVPTVEENIAGAGSALNTFLQAGVLTMPVNCIGLDPDFVEQSGREBLISVG 605
Db 551 LAGSHELLVLTIEENVMVGAGSAGVEFLASEGLLEVPDLGLDIDYVYHAKPSEMLAEQG 610
Qy 606 LPSKGLATIEQ 617
Db 611 LDAQIGERAVRQ 622

```

RESULT 7
AAU00013
ID AAU00013 standard; Protein; 627 AA.

```

XX AC AAU00013;
XX DT 08-MAY-2001 (first entry)
XX DE Pseudomonas aeruginosa dxs ts-32 mutant.
XX KW Deoxyxylose-5-phosphate synthase; dxs; agonist; antagonist;
XX KW diagnostic; eukaryotic extracellular matrix protein; antibody; wound;
XX KW in-dwelling device; Helicobacter pylori infection; gastric ulcer;
XX KW gastritis; gastrointestinal carcinoma; mutant; mutagen; ts-32.
XX OS Pseudomonas aeruginosa.
XX FH Key Location/Qualifiers
XX FT Misc-difference 115 /note= "Wild-type Pro replaced by Ser"
XX FN MO200114409-A1.
XX PD 01-MAR-2001.
XX PE 16-AUG-2000; 2000QMO-US22332.
XX PR 19-AUG-1999; 99US-0377279.
XX PA (SMIX ) SMITHKLINE BEECHAM CORP.
XX PA (SMIX ) SMITHKLINE BEECHAM PLC.
XX PI Huang J, Jiang X, McDevitt D;

```

```

XX WP1: 2001-211303/21.
DR N-PSDB; AAS00018.
XX PT New deoxyxylose-5-phosphate (dxs) polypeptides of Pseudomonas
XX PT aeruginosa useful for diagnosis of dxs expression or activity related
XX PT disease -
XX PS Example 2; Page - ; 41pp; English.
XX CC The sequence represents Pseudomonas aeruginosa deoxyxylose-5-phosphate
CC synthase, dxs, temperature sensitive mutant, ts-32. (Ant)agonists of dxs
CC are useful for treating an individual in need of enhanced or inhibited
CC activity or expression of or immunological response to dxs. Dxs and the
CC nucleic acid encoding it are useful as diagnostic reagents, especially in
CC relation to diseases related to expression or activity of dxs. Fragments
CC of the nucleic acid are useful as probes or primers and to synthesize
CC full length dxs polynucleotides. The nucleic acid and protein are useful
CC for diagnosing bacterial infections, especially by Pseudomonas
CC aeruginosa, staging a disease or response of an infectious organism to
CC drugs, and are also used in a structure based design of an (ant)agonist
CC to dxs. The nucleic acid, protein and antibodies raised against it are
CC useful for screening for (ant)agonists of dxs or the nucleic acid and are
CC pathogen and a mammalian host responsible for sequelae of infection.
CC These molecules are useful in preventing adhesion of gram positive and/or
CC gram negative bacteria to eukaryotic extracellular matrix proteins or
CC in-dwelling devices or in wounds, to block bacterial adhesion; to block
CC the normal progression of pathogenesis in infections initiated other than
CC by the implantation of in-dwelling devices or by other surgical
CC techniques. Dxs is also useful for identifying membrane bound or soluble
CC receptors. The (ant)agonists are useful for treating Helicobacter pylori
CC infection, and also to prevent, inhibit or treat gastric ulcers,
CC gastritis and gastrointestinal carcinoma.
CC The present sequence is not shown in the specification but is derived
CC from the dxs sequence given in AAU00012.
XX SQ Sequence 627 AA;

```

Query Match 64.1%; Score 2061.5; DB 22; Length 627;
Best Local Similarity 63.2%; Pred. No. 9.3e-180;
Matches 387; Conservative 100; Mismatches 118; Indels 7; Gaps 2;

```

Qy 8 PLKNIHTPADIALSKDQLOGLADREKGYLTHTVTSISGCHFAAGLCTVLTVAHLYFN 67
Db 16 PLDRASSPAELRRLTGADLETLADLRQYLTIVGQGTGHFGAGLVGLTTLAHLVFD 75
Qy 68 TPVDQLVMDVGHQAVPHKILTERBELMGLRQKNGLAAPRRAESEYDFRGHSTSTIS 127
Db 76 TPDDRLVMDVGHQAVPHKILTERBELMGLRQKNGLAAPRRAESEYDFRGHSTSTIS 135
Qy 128 AALGMAIASQLGEGDKKKVAIIGDSITGMAAYEAMNAGDVANLIVLINDMSISPP 187
Db 136 AALGMAIAARLOQEKERSVAIVGDGALTAGMAFEALNHASEVDADMEVLINDMSISHN 195
Qy 188 VGAMNNYLTKVLSKREYSSVREESKKALAMPVWEIARKEEYKGMIVPGTLFEEIGF 247
Db 196 VGGLSNTYLAKLISRRYSRMSRSGSKVLSRLPGAMBIARTETEVAKKGLVPTLLEELGM 255
Qy 248 NYFGPIDGHVEMLVSTENLKDITGVPVLEVVYTKKGAVAPAEKDPPLAYHGVPAFDPK 307
Db 256 NYIGPIDGHDLPTLVATLNNRDMKGPQFLHVVTKKGKGFAPALDPIGYHAI-----TK 310
Qy 308 DFLPKAAP--SPHPTTVYVGRMLCDMAADERLIGITPAMEGSGLVESGKZPNRYPD 365
Db 311 LEAPGSAFKKGTGPKTSVFGQMLCDMAADARLLGITPAMEGSDLVAFSERYPERYFD 370
Qy 366 VALAQHAAVTLAAGACQAGAKPVAIYSTFLQRYDOLIHVVALQNLDMPLALDRAGLVG 425
Db 371 VALAQHAAVTLAAGACQAGAKPVAIYSTFLQRAYDOLIHVAVQHLDLVFAIDRAGLVG 430
Qy 426 PGGPTTAGAFDYSYKACIPNMLIMAPADNECQMLTTFQFHGPASTRYPRGKPGALAI 485

```

Db 431 EDGPTAGSFDISYACIPQMLVMPSPDDEDEKLKLTGVLFDGPAVAVPRGSGPNHET 490
 QY 486 DPLTLALTEGKAEVRHGHRIALANGSVTPAVEKQKQATVVMRRVPPDQALVLE 545
 Db 491 DPLQVEVEIGKVGVRGGVALVFGVLAEMKVAESLDAVVMRRVPPDQALVLE 550
 QY 546 LARTHEVFTVBENAVAGASAINFTFLOKQKMLPVCNIGLPDRVEGSGREBELSLVG 605
 Db 551 LAGSHLAVTIEBNAMGAGSANGVGFASGLAEVFLQGLPDTYVEAKSSEMLABCG 610
 QY 606 LDSKGIATIEQ 617
 Db 611 LDAAGIEKAVRQ 622
 RESULT 8
 ID AAU38490 standard; Protein; 620 AA.
 XX AAU38490;
 XX 14-FEB-2002 (first entry)
 DT Salmonella typhi cellular proliferation protein #381.
 DE Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 OS Salmonella typhi.
 XX WO200170955-A2.
 XX 27-SEP-2001.
 PD 21-MAR-2001; 2001WO-US09180.
 PF 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 PA (ELIT-) ELITRA PHARM INC.
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 DR WPL; 2001-611495/70.
 DR N-PSDB; AAS56349.
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX Example 3; Seq ID No 14083; 511pp; English.
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX
 SQ Sequence 620 AA;
 Query Match 62.6%; Score 2014.5; DA 22; Length 620;
 Best Local Similarity 61.0%; Pred No. 1.9e-175;
 Matches 375; Conservative 106; Mismatches 131; Indels 3; Gaps 3;
 QY 7 YELKNHTPADIPALSKDQLOLADEVGYLTHTVSISGGHFAAGTVELVLAHVAF 66
 Db 8 YPTLAVSTQELRLPRESLPKLCDELRVLDSVSSSGHFAAGTVELVLAHVAF 67
 QY 67 NTPVDQLVWDVGHQAVPKKILTKRKEKPTRTIGGVSAFPARDBSEPDAGVGHSSSTSI 126
 Db 68 NTPFDQLIWDVGHQAVPKKILTKRRDKGTTRQGGHLPFPWRGSEPDVLSVGHSSSTSI 127
 QY 127 SAALGMATASOLRGECKQVAIIIGDSITGGMAYEAMNHADVNALEVIINDNDMSISP 186
 Db 128 SAGIGIAVAEKEGKDRRTVCVIGGALTAGMAPAMNHADIRPDMVILINDNDMSISB 187
 QY 187 PVGAMNNTLYVLSKKEYSVVBESKKAALAMPVSWELARTTEHKVAKTVPGLFEEIG 246
 Db 188 NVGALNNHIALRLSGKLYSSIREGKRVSGVPPIKELKTEEHIKGVVPGTLFEEIG 247
 QY 247 FNYEGPIDGHVEMVSTLENLKDITGVPFLVVTKKKGAPAPAKDPLAVHGVAPFPT 306
 Db 248 FNYIGPVGHVDMGLISTLKNRRDKGQPLMTKRGVGPAPKDDPTTHAVKPDPS 307
 QY 307 KDFLEKAPSPHPTTYTEVFGMLCDMAODERLLGITPAMEGSGVFSQKFPNRYPDV 366
 Db 308 SGCLPKSS-GGLPGYSKIFGDMLCETAKDSKLAITPAMEGSGVFSKRPFRYPDV 366
 QY 367 AIAEGHATTLAAGACQCAKPVVAIYSTFLORGPDOLIHDVALQNLMLFALDPAGLVGP 426
 Db 367 AIAEGHATVFAAGIAGGVKPVVAIYSTFLORADVHDVALQNLMLFALDPAGLVGP 426
 QY 427 DGPTTAGAFDYVYRCIEPNMLINAPADENECRQMLTGGFQHH-GPASYVRGKGPAGAI 485
 Db 427 DGQTHGAFDLSYACIPQMLVMPSPDDEDEKLKLTGVLFDGPAVAVPRGSGPNHET 486
 QY 486 DPLTLALTEGKAEVRHGHRIALANGSVTPAVEKQKQATVVMRRVPPDQALVLE 545
 Db 487 TP-LEKLPITGKGLVKGHEKLAIIIFGTLMEPAKVALEALNATVDMKFPVPLDITLIE 545
 QY 546 LARTHEVFTVBENAVAGASAINFTFLOKQKMLPVCNIGLPDRVEGSGREBELSLVG 605
 Db 546 MAQCHDALVTLSENALMGAGSANGVGFASGLAEVFLQGLPDTYVEAKSSEMLABCG 605
 QY 606 LDSKGIATIEQ 620
 Db 606 LDAAGIEKAVRQ 620
 RESULT 9
 ID AA52832 standard; Protein; 620 AA.
 XX AA52832;
 XX 26-JAN-2000 (first entry)
 DT *Escherichia coli* protein sequence SEQ ID NO:1.
 DE *Escherichia coli*
 XX Isoprenoid; microorganism; detection; antibacterial; herbicide;
 KM heart disease; osteoporosis; haemostasis; cancer; immunopotential;
 KM healthy food; antifouling coating; farnesyl pyrolonic acid; pyruvic acid;
 KM 1-deoxy-D-xylose-5-phosphate; glyceraldehyde-3-phosphate;
 KM 2-C-methyl-D-erythreitol-4-phosphate.
 OS *Escherichia coli*.

PN WO953071-A1.
 XX
 PD 21-OCT-1999.
 XX
 PF 14-APR-1999; 99WO-0P01987.
 XX
 PR 14-APR-1998; 98JP-0103101.
 PR 05-AUG-1998; 98JP-0221910.
 PR 15-FEB-1999; 99JP-0035739.
 XX
 PA (KYO) KYOWA HAKKO KOGYO KK.
 PI Miyake K, Hashimoto S, Motoyama H, Ozaki A, Seto H, Kuzuyama T,
 PI Takahashi S;
 XX WPI: 1999-620434/53.
 DR N-PSDB; AA233157.
 XX
 PT Preparation of recombinant isoprenoid compounds useful for treatment of
 PT heart diseases, osteoporosis and hemostasis, preventing cancer and
 PT immunopotentialiation
 XX
 PS Claim 2; Page 63-67; 145pp; Japanese.
 CC The present invention describes the preparation of an isoprenoid
 CC compound comprising using at least 1 DNA e.g. encoding proteins which
 CC elevate the efficiency of the synthesis or DNA encoding a fattyacyl
 CC pyrolytic producing enzyme. The method of preparation of an isoprenoid
 CC compound comprises using at least 1 DNA, a vector, cloned cells, their
 CC derived recombinant DNAs or transformed products in a culture system and
 CC extracting the isoprenoid accumulated in the medium. The DNA encodes at
 CC least 1 of the following: (a) a compound for activating or catalyzing
 CC the production of 1-deoxy-D-xylulose-5-phosphate from pyruvic acid and
 CC glyceraldehyde-3-phosphate; (b) an enzyme producing fattyacyl pyrolytic
 CC acid; (c) a protein which elevates the efficiency of synthesis of
 CC isoprenoid compounds and comprises a 3 or 4 amino acid sequence
 CC optionally with 1 or more of the amino acids being deleted or
 CC substituted or an additional amino acid being inserted; (d) a protein
 CC which activates or catalyzes the production of 2-C-methyl-D-erythritol-
 CC 4-phosphate from 1-deoxy-D-xylulose-5-phosphate; or (e) a protein which
 CC activates a target compound or reaction and is a string and or hybrid of
 CC the DNA encoded in (a)-(e). Isoprenoid compounds are useful in drugs
 CC (e.g. for the treatment of heart diseases, osteoporosis and hemostasis,
 CC for preventing cancer and as immunopotentializers), health foods and
 CC antifouling coatings. The isoprenoids also inhibit enzymatic reactions
 CC on the non-mevalonate pathway and can be used as antibacterials and
 CC herbicides. The present sequence is used in the exemplification of
 CC the present invention.
 CC
 SQ Sequence 620 AA;
 Query Match 62.2%; Score 2000.5; DB 20; Length 620;
 Best Local Similarity 60.2%; Pred. No. 3.6e-174;
 Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;
 QY 7 YPLIKIHHPADIRALSKQLOQLADENVGYLTHTYSGSGFPAAGISGVETVALHYVF 66
 DB 8 YPTALVDTSTQELRLPKESLPQLCDLRLRYLIDSSRSRSGFASGLVETVALHYVF 67
 QY 67 NTFVDDLVWDVGHQAVPAKILTGKREKMPRTIRLTGSVSAFPADESEYDAFGVGHSTSI 126
 DB 68 NTFPDLQIDVGHQAVPAKILTGKREKMPRTIRLTGSVSAFPADESEYDAFGVGHSTSI 127
 QY 127 SAALGMAISQLRGEKQKVAITIGDSITGMAFYEMNHAGDVNANLITINDNMSISP 186
 DB 128 SAIGIGAVAAEKRGKRRITVCYIGDAITPAGMAFEMNHAGDI RPDMLVINDNMSISE 187
 QY 187 PGCAANNVITKLVSSKESVSESKKALAKMSVWELARKTEBHVKGAVIGTIFEEIG 246
 DB 188 NFGALNNHLLAQLLSGLYSSLRGKXVSGVPIKELKREBHKKGAVGTIFEEIG 247
 QY 247 ENYEGSIDGHVEMLVSTLENKQLTGPPVFLAVVTYKKGAVPAEADPLATGVPAFDP 306

DB 248 ENYIGPVGDHVLGLITLLNMMDLKGPQLHMTKKRGVEPAEKDPTIFHAVKFPDS 307
 QY 307 KDFPFAAPSPHPITTYVFGRLCDMAAQBRLIGITPARESGSLVEPSQKFNRRFPDV 366
 DB 308 SGCLPKSS-GGLPYSKIPDWMLCETAQKDKMALITPARESGSWEFRRKFPDRFPDV 366
 QY 367 AIAEQHAVTLAAGACGAKPVVAIVYSTFLORGYDQLIHVALQNLMLPALDRAGLVGP 426
 DB 367 AIAEQHAVTPAAGALIGYPIVAIVYSTFLORADYDLHVALQKLVLRALDAGLVGA 426
 QY 427 DGPTHAAGFDYSYKRCIPNMLINAPADENECRQMLTTGPOH-CPASVRYPRGPGAAI 485
 DB 427 DGQTHGAFDLSYLRCLPEWIMTPSDNECRQMLYGYHYNDPSAVRYPRGPAVGL 486
 QY 486 DPTTALIEIGAEVHHGSRITAIAMGSMTPPAVEAKOQCATVNNRFPVGPPOALVLE 545
 DB 487 TP-LEKLPICKGIVKRGKEKALINFTITLPEAKVAESLNAITVDRFVYKPDLEALILB 545
 QY 546 IARTHDVFTVYENVIAGAGSALNTEFLQAKVIMPCNIGLPDRFVYEGSREELSLVG 605
 DB 546 MAASHMALVEYENALINGAGSGVNEVIMARRKVPVLTIGLPFFIPOGTQEMERAILG 605
 QY 606 LDSGILATIEQFCA 620
 DB 606 LDAAGMEAKIKAWLA 620
 RESULT 10
 ID AAY08880 standard; protein; 620 AA.
 AC AAY08880;
 XX 13-AUG-1999 (first entry)
 DE E. coli DXS protein.
 XX 1-deoxy-D-xylulose-5-phosphate synthase; DXS; modulator; pyruvate;
 KM glyceraldhyde-3-phosphate; GA3P; 1-deoxy-D-xylulose-5-phosphate; DXP;
 KM structural analogue; pesticide; antibacterial; herbicide; biosynthesis;
 KM growth promoter; DXP derivative; isoprenoid; carotenoid; chlorophyll;
 KM phytol; lutein; sterol; ubiquinone; rubber; taxane.
 OS Escherichia coli.
 XX DE29800547-U1.
 XX 08-APR-1999.
 XX 16-JAN-1998; 98DB-2000547.
 XX 28-NOV-1997; 97DB-1052700.
 PA (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
 PA (AGRE) HOECHST-SCHERING AGRERO GMBH.
 XX WPI: 1999-230818/20.
 DR 1-deoxy-D-xylulose-5-phosphate synthase protein and modulators
 PT potentially useful as pesticides, antibacterial agents, herbicides,
 PT etc.
 PS Claim 3; Page 10; 21pp; German.
 CC This invention describes a 1-deoxy-D-xylulose-5-phosphate synthase (DXS)
 CC protein or active fragment and a DXS modulator. The DXS modulator is a
 CC structural analogue of pyruvate, glyceraldehyde-3-phosphate (GA3P) or
 CC 1-deoxy-D-xylulose-5-phosphate (DXP). Compounds that inhibit DXS activity
 CC are potentially useful as pesticides, antibacterial agents or herbicides.
 CC Compounds that stimulate DXS activity are potentially useful as growth
 CC promoters or for increasing biosynthesis of DXP derivatives, especially
 CC isoprenoids such as carotenoids, chlorophyll, phytol, lutein, sterols,
 CC ubiquinone, rubber and taxanes.

QY 606 LDSKGLATIPQFCA 620
 DB 606 LDAAGEAKIKAWLA 620

RESULT 12

AAV51613
 ID AAV51613 standard; Protein; 620 AA.

AAV51613:

26-MAY-2000 (first entry)

E. coli DOXS protein.

DOXS; 1-deoxy-D-xylulose-5-phosphate synthase; Hppp; GGPPOR; plant;
 p-hydroxyphenylpyruvate dioxygenase; tocopherol; vitamin K; chlorophyll;
 geranylgeranyl-pyrophosphate oxidoreductase; carotenoid; transgenic; ss.
 Escherichia coli.

MO200008169-A1.

17-FEB-2000.

30-JUL-1999; 99MO-BP05467.

05-AUG-1998; 98DE-1035219.

01-OCT-1998; 98DE-1045216.

01-OCT-1998; 98DE-1045224.

01-OCT-1998; 98DE-1045231.

(SUNG-) SUNGENE GMBH & CO KGAA.

Reindl A, Mejia PL, Palmas JME, Gracia MAC, Ebneith M, Herbers K;

WPI: 2000-195578/17.

N-PSDB; AA288977.

Use of DNA encoding 1-deoxy-D-xylulose-5-phosphate synthase to produce

plants with increased tocopherol, vitamin K, chlorophyll and carotenoid

content

Example 3; Page 76-78; 94pp; German.

This invention describes the novel use of a DNA sequence encoding

1-deoxy-D-xylulose-5-phosphate synthase (DOXS), and optionally

p-hydroxyphenylpyruvate dioxygenase (Hppp) and/or geranylgeranyl-

pyrophosphate oxidoreductase (GGPPOR), to produce a plant with increased

tocopherol, vitamin K, chlorophyll and/or carotenoid content. Transgenic

plants containing DOXS DNA coding sequences can be used for production of

plant and bacterial DOXS which also have increased tocopherol, vitamin K,

chlorophyll and/or carotenoid content. The test system can be used to

identify inhibitors of DOXS. This sequence represents the Escherichia

coli DOXS protein described in the method of the invention.

Sequence 620 AA:

Query Match 62.2%; Score 2000.5; DB 21; Length 620;
 Best Local Similarity 60.2%; Pred. No. 3.6e-174;
 Matches 370; Conservative 114; Mismatches 126; Indels 3; Gaps 3;

QY 187 PVGANNVLTUVSSKEFYSSREBSKKALAMPSEVWEIARTEEHVKGMTVPGTLPEETG 246
 DB 188 NVGALNNHIAOLLSGKLYSSIREGKKVFGCVPIKXILKRTETIKGAVVPGTLPEETG 247
 QY 247 FNYFGPIDGHVEMVSTLBNKDLTGPVPLHVVYTKKGGYAPAKDPLAHGVGPARDPT 306
 DB 248 FNYIGPVVDHVLGLITLKNMRDLKGPQLHMTTKXGRGYEPABKOPITFHAVPKFDP 307
 QY 307 KDLPLKAPSPHPTVTEVFGMLCDMAAODERLLGITPARREGSGLVESFQKPEPRYFDV 366
 DB 308 SGCLPKRS-GGLPSYSKIPGDMLCETAKONKMAITPAMEGSGMVEFSRKFPRYFDV 366
 QY 367 AIAEGHAYTLAAGACQAKVVALVYSTPLOGYDOLIHVALNMLPFLDPAAGLVGP 426
 DB 367 AIAEGHAYTLAAGACQAKVVALVYSTPLOGYDOLIHVALNMLPFLDPAAGLVGP 426
 QY 427 DGPTRGAPVDSYNNRCPNMLIMPADENECRQMLTGPQHH-GRASVRYRGGPGGAI 485
 DB 427 DGPTRGAPVDSYNNRCPNMLIMPADENECRQMLTGPQHH-GRASVRYRGGPGGAI 485
 QY 486 DPTLTALFEGKAFVRRHGSRIAILAMGSMVTPAVEAGKOLGATVNNRFPVPPDQATLE 545
 DB 487 TP-LEKLPISGKIVRRRGEKALINPGLMPKPAKVAESLNAITVDMKFTVAPDLBALILE 545
 QY 546 LARTHDVFTVEENVVAGAGSAINTEFLOAKVLMPCNIGLPDRFVQGSREELSLVIG 605
 DB 546 MAASHBALVTEENAVINAGAGSGVNEVLMARKEVPLNIGLPDRFVQGSREELSLVIG 605
 QY 606 LDSKGLATIPQFCA 620
 DB 606 LDAAGEAKIKAWLA 620

RESULT 13

AAU34475
 ID AAU34475 standard; Protein; 620 AA.

AAU34475:

14-FEB-2002 (first entry)

E. coli cellular proliferation protein #56.

Antisense; prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

Escherichia coli.

MO200170955-A2.

27-SEP-2001.

21-MAR-2001; 2001MO-US09180.

21-MAR-2000; 2000US-191078P.

23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-207727P.

23-OCT-2000; 2000US-242578P.

22-NOV-2000; 2000US-253625P.

22-DEC-2000; 2000US-257931P.

16-FEB-2001; 2001US-269308P.

(ELIT-) ELITRA PHARM INC.

Haseelbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

Yamamoto RT, Xu HH;

WPI: 2001-611495/70.

N-PSDB; AAS52334.

New polynucleotides for the identification and development of

antibiotics, comprise sequences of antisense nucleic acids -

Example 3; Seq ID No 10068; 511pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_pct_sequences.

Sequence 620 AA:
Query Match 62.2%; Score 2000.5; DB 22; Length 620;
Best Local Similarity 60.2%; Pred. No. 3.6e-174;
Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

7 YPLKNIHTPADIRALSDQLOGLADVRGTLHTVSISGGHAAAGTVELTALHYVP 66
8 YPLALVDSIOELRLPKESLPKLCDELRLYLDVSRSRSGHPSAGTVELTALHYV 67
67 NTPVDQVMDVGHQAYPHKILITGRKEMPTIRLGVSAPFARDESEYDAFGVGSSTSI 126
68 NTPFDQIMDVGHQAYPHKILITGRKEMPTIRLGVSAPFARDESEYDAFGVGSSTSI 127
127 SAALGMALASQLRGDEKRVVAIIIGDSITGGMAYEAMNHAGDVANLVILINDNMSISP 186
128 SAGIGIAVAAREKGNRRVCVIGDGLTAGAFAEMNHAGDIRPMLVILINDNMSISB 187
187 PVGAMNNYITLVSSKPYSSVRESKKALAKMPSVWEIARKTEEHKGVIVPGLTFFELG 246
188 NVGALNNHLAGLISGLYSLRGGKKVSGVPPIKELIKRTEBHKGVVPGTLFFELG 247
247 FNYFGPIDGHVEMLVSTLENKDLTGVPVPLAVVTKKGVAPAEKDPPLAVGVPAPDPT 306
248 FNYIGPVGDHVLGLITLTKNMRDLKGPQFLIMTKGSGYEPAEKDPITFHAVPKDPS 307
307 KDFLPAAASHRTTYEVFGRMLCMAAODERLIGTPMREGSGIVERSOKFPNRYPPV 366
308 SGCLPKRS-S-GGLPSYSKIRGDM/CETPAADNKLMAITTPMREGSGIVERSOKFPNRYPPV 366
367 ALAEGAVVTLAAGQAQGAAPVAIYSTFLORGYPDILHVALONLMDLFAIDRAGLVGP 426
367 ALAEGAVVTLAAGQAQGAAPVAIYSTFLORGYPDILHVALONLMDLFAIDRAGLVGP 426
427 DQGTGGAADLSTLRICIPENVIMTPSDENECROMLYTGVIYNDGPAVAVPRGNAVGVEL 485
427 DQGTGGAADLSTLRICIPENVIMTPSDENECROMLYTGVIYNDGPAVAVPRGNAVGVEL 485
486 DPLTALTEIGKAEVRRHSGRIALANGSVMTPAVEGKQLGATVUNMRVKKPDALVTE 545
487 TP-LEKLPKIGKIGVKGRCETALINGTMLPAPAKAESLNTLVDMRVRKPLDELTIE 545
546 LARTHDVFTVEENIAGAGSAINTFLOAKVLMPCNIGLIDRFEVGGSEEBELSLVG 605
546 MAASHBALVTVEENIAGAGSAINTFLOAKVLMPCNIGLIDRFEVGGSEEBELSLVG 605
606 LSKGILATIEQCA 620
606 LSKGILATIEQCA 620
DB 606 LSKGILATIEQCA 620

RESULT 14

ID AA021860 standard; Protein; 620 AA.

AA021860;

13-SEP-2002 (first entry)

Isoprenoid related protein sequence SEQ ID No 31.

Isoprenoid: CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DKS;

decaprenyl diphosphate synthase.

Escherichia coli.

WO200226933-A2.

04-APR-2002.

28-SEP-2001; 2001WO-US30328.

29-SEP-2000; 2000US-236580P.

(CRGI) CARGILL INC.

Gokarn R, Jessen H, Zidwick MJ;

WPI; 2002-416480/44.

Substantially pure polypeptides having e.g., 1-deoxyxylulose-5-phosphate synthase activity, useful for the production of isoprenoids, especially CoQ(10).

Disclosure; Fig 6; 246pp; English.

The invention relates to methods and materials for the production of isoprenoids. More particularly the invention provides isolated nucleic acids, substantially pure polypeptides, host cells, and methods for producing various isoprenoid compounds. The polypeptides are useful for the production of isoprenoids, especially CoQ(10). Expressing the pure polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS) activity or decaprenyl diphosphate synthase (DDS) activity, is useful for increasing production of CoQ(10) in a cell having endogenous DDS activity. This sequence represents a protein relating to the isoprenoid production of the invention.

Sequence 620 AA:
Query Match 62.2%; Score 2000.5; DB 23; Length 620;
Best Local Similarity 60.2%; Pred. No. 3.6e-174;
Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

7 YPLKNIHTPADIRALSDQLOGLADVRGTLHTVSISGGHAAAGTVELTALHYVP 66
8 YPLALVDSIOELRLPKESLPKLCDELRLYLDVSRSRSGHPSAGTVELTALHYV 67
67 NTPVDQVMDVGHQAYPHKILITGRKEMPTIRLGVSAPFARDESEYDAFGVGSSTSI 126
68 NTPFDQIMDVGHQAYPHKILITGRKEMPTIRLGVSAPFARDESEYDAFGVGSSTSI 127
127 SAALGMALASQLRGDEKRVVAIIIGDSITGGMAYEAMNHAGDVANLVILINDNMSISP 186
128 SAGIGIAVAAREKGNRRVCVIGDGLTAGAFAEMNHAGDIRPMLVILINDNMSISB 187
187 PVGAMNNYITLVSSKPYSSVRESKKALAKMPSVWEIARKTEEHKGVIVPGLTFFELG 246
188 NVGALNNHLAGLISGLYSLRGGKKVSGVPPIKELIKRTEBHKGVVPGTLFFELG 247
247 FNYFGPIDGHVEMLVSTLENKDLTGVPVPLAVVTKKGVAPAEKDPPLAVGVPAPDPT 306
248 FNYIGPVGDHVLGLITLTKNMRDLKGPQFLIMTKGSGYEPAEKDPITFHAVPKDPS 307

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 ; Search time 42.3225 Seconds
(without alignments)
3044.503 Million cell updates/sec

Title: US-09-941-947a-6

Perfect score: 3216

Sequence: 1 MKLITDYPILKNIHTPADIR.....LSIVGDSKGIATTEQFCA 620

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3216	100.0	620	US-09-934-903-2	Sequence 2, Appl1
2	3216	100.0	620	US-09-934-868-62	Sequence 62, Appl1
3	3216	100.0	620	US-09-941-947a-6	Sequence 6, Appl1
4	2080	64.7	628	US-10-369-493-13998	Sequence 13998, A
5	2069.5	64.4	627	US-10-381-779-15	Sequence 15, Appl1
6	2014.5	62.6	620	US-09-815-242-14083	Sequence 14083, A
7	2004.5	62.3	619	US-10-369-493-481	Sequence 481, A
8	2003.5	62.3	620	US-10-369-493-21174	Sequence 21174, A
9	2000.5	62.2	620	US-09-815-242-10068	Sequence 10068, A
10	2000.5	62.2	620	US-10-381-779-31	Sequence 31, Appl1
11	2000.5	62.2	620	US-10-369-493-731	Sequence 731, Appl1
12	1992.5	62.0	623	US-10-369-493-15911	Sequence 15911, A
13	1992.5	62.0	623	US-10-369-493-16281	Sequence 16281, A
14	1992.5	62.0	653	US-10-369-493-15540	Sequence 15540, A
15	1990.5	61.9	619	US-10-369-493-7428	Sequence 7428, Ap

16	1985.5	61.7	616	US-10-369-493-4669	Sequence 4669, Ap
17	1971	61.3	615	US-10-369-493-9272	Sequence 9272, Ap
18	1969.5	61.2	618	US-10-369-493-8507	Sequence 8507, Ap
19	1967	61.2	613	US-10-369-493-9367	Sequence 9367, Ap
20	1951	60.7	670	US-10-369-493-17644	Sequence 17644, A
21	1937.5	60.2	625	US-10-381-779-33	Sequence 33, Appl1
22	1776	55.2	637	US-10-381-779-32	Sequence 32, Appl1
23	1623	50.5	624	US-10-369-493-10188	Sequence 10188, A
24	1620.5	50.4	628	US-10-381-779-18	Sequence 18, Appl1
25	1620.5	50.4	628	US-10-369-493-61	Sequence 61, Appl1
26	1555	48.4	630	US-10-369-493-20596	Sequence 20596, A
27	1543.5	48.0	637	US-10-369-493-12290	Sequence 12290, A
28	1532.5	47.7	641	US-10-381-779-3	Sequence 3, Appl1
29	1530.5	47.6	629	US-10-369-493-10848	Sequence 10848, A
30	1523.5	47.4	640	US-10-369-493-10842	Sequence 10842, A
31	1503.5	46.8	654	US-10-369-493-7893	Sequence 7893, Ap
32	1499.5	46.6	630	US-10-369-493-14775	Sequence 14775, A
33	1499.5	46.6	634	US-10-369-493-11753	Sequence 11753, A
34	1499.5	46.6	635	US-10-369-493-14237	Sequence 14237, A
35	1499.5	46.6	635	US-10-369-493-14928	Sequence 14928, A
36	1491.5	46.4	648	US-10-381-779-27	Sequence 27, Appl1
37	1485.5	46.2	720	US-10-427-590-26	Sequence 26, Appl1
38	1482	46.1	625	US-10-369-493-18033	Sequence 18033, A
39	1468	45.6	614	US-10-369-493-10317	Sequence 10317, A
40	1459	45.4	629	US-10-369-493-17337	Sequence 17337, A
41	1444	44.9	635	US-10-369-493-20015	Sequence 20015, A
42	1444	44.9	671	US-10-259-194A-414	Sequence 414, Appl1
43	1442	44.8	640	US-10-381-779-29	Sequence 29, Appl1
44	1442	44.8	640	US-10-369-493-2645	Sequence 2645, Ap
45	1439	44.7	636	US-10-381-779-28	Sequence 28, Appl1

ALIGNMENTS

RESULT 1
US-09-934-903-2
; Sequence 2, Application US/09934903
; Patent No. US20020102690A1
GENERAL INFORMATION:
APPLICANT: Koffas, Mattheos
APPLICANT: Odum, J. Martin
APPLICANT: Schenzle, Andreas J.
APPLICANT: No. US20020102690ALton, Kelley C.
APPLICANT: Tomb, Jean-Francois
APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
FILE REFERENCE: C11646 US NA
CURRENT APPLICATION NUMBER: US/09/934, 903
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/229,907
PRIOR FILING DATE: September 1, 2001
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 2
LENGTH: 620
TYPE: PRT
ORGANISM: Methylobionas 16a
FEATURES:
OTHER INFORMATION: Amino acid sequences encoded by ORF1
US-09-934-903-2

Query Match 100.0%; Score 3216; DB 10; Length 620;
Best Local Similarity 100.0%; Pred. No. 1.1e-298;
Matches 620; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLITDYPILKNIHTPADIRALSDQOQLADYRGVLTIVTSISGSHFAGLTVTLV 60
DB 1 MKLITDYPILKNIHTPADIRALSDQOQLADYRGVLTIVTSISGSHFAGLTVTLV 60
QY 61 ALHVVETPVDOILWVDVGHQAPHPKILTKRKRPMTIRTLTGVSAPFADESEYDAGVG 120

Db 61 ALHYVFNTPVDQLVWDVGHQAYPHKILTGKREMPITRTLGVSAPPADBSFYDAFVG 120
 QY 121 HSSSTISALGMAIALSOLRGEDKRWVAITGGSGITGMAVEAMNHAGVNNMLVTILND 180
 Db 121 HSSSTISALGMAIALSOLRGEDKRWVAITGGSGITGMAVEAMNHAGVNNMLVTILND 180
 QY 181 DMSISPPVGAMNNYLTKVLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPGT 240
 Db 181 DMSISPPVGAMNNYLTKVLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPGT 240
 QY 241 LPEELGFNFGPDIHGDMVEMVSTLENLKDITGVPVFLHVTYKKGVAAPAKDPLAHGV 300
 Db 241 LPEELGFNFGPDIHGDMVEMVSTLENLKDITGVPVFLHVTYKKGVAAPAKDPLAHGV 300
 QY 301 PAFDPTKDFLPAKAPSPPHTYTEVFGRMLCDMAADERLLGITPAMREGSGLVEFSQKFP 360
 Db 301 PAFDPTKDFLPAKAPSPPHTYTEVFGRMLCDMAADERLLGITPAMREGSGLVEFSQKFP 360
 QY 361 NRYFDVAIAEQHAAVTLAAGACQCAKPVVAIYSTFLQRYDQILHDVALQNLDMFLALDR 420
 Db 361 NRYFDVAIAEQHAAVTLAAGACQCAKPVVAIYSTFLQRYDQILHDVALQNLDMFLALDR 420
 QY 421 AGLVGDPGPTHAGAFDYSYMRCLPNNLIMAPADNECRQMLTTGPOHGPASVRYPRGKG 480
 Db 421 AGLVGDPGPTHAGAFDYSYMRCLPNNLIMAPADNECRQMLTTGPOHGPASVRYPRGKG 480
 QY 481 PGAADPTLTALBIGAEVRHSGRIAILAMGSMVTPAVEAGKOLGATVNNRFPVKPDQ 540
 Db 481 PGAADPTLTALBIGAEVRHSGRIAILAMGSMVTPAVEAGKOLGATVNNRFPVKPDQ 540
 QY 541 ALVIELARTHDVPTVVEENVVIAAGAGSAINTEFLQAKVMPVCNIGLPDRFVEQSSREL 600
 Db 541 ALVIELARTHDVPTVVEENVVIAAGAGSAINTEFLQAKVMPVCNIGLPDRFVEQSSREL 600
 QY 601 LSLVGLDSKGLIATIEQFCA 620
 Db 601 LSLVGLDSKGLIATIEQFCA 620

RESULT 2 US-09-934-868-62

; Sequence 62, Application US/09934868
 ; Patent No. US2002037190A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kofas, Matcheos
 ; APPLICANT: Odum, James M.
 ; APPLICANT: Schenke, Andreas J.
 ; TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN
 ; FILE REFERENCE: CL1596 US NA
 ; CURRENT APPLICATION NUMBER: US/09/934,868
 ; CURRENT FILING DATE: 2001-08-22
 ; PRIOR APPLICATION NUMBER: 60/229,858
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 81
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 62
 ; LENGTH: 620
 ; TYPE: PRT
 ; ORGANISM: Methylomonas 16a
 ; FEATURE:
 ; OTHER INFORMATION: Amino acid sequences encoded by DXS
 US-09-934-868-62

Query Match 100.0%; Score 3216; DB 10; Length 620;
 Best Local Similarity 100.0%; Pred. No. 1,le-298;
 Matches 620; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLTIDYPLKNIHTPADIRALSKDQLQQLADEVGVLTHTVSISGGHFAAGLGTVELTV 60
 Db 1 MGLTIDYPLKNIHTPADIRALSKDQLQQLADEVGVLTHTVSISGGHFAAGLGTVELTV 60
 QY 61 ALHYVFNTPVDQLVWDVGHQAYPHKILTGKREMPITRTLGVSAPPADBSFYDAFVG 120

Db 61 ALHYVFNTPVDQLVWDVGHQAYPHKILTGKREMPITRTLGVSAPPADBSFYDAFVG 120
 QY 121 HSSSTISALGMAIALSOLRGEDKRWVAITGGSGITGMAVEAMNHAGVNNMLVTILND 180
 Db 121 HSSSTISALGMAIALSOLRGEDKRWVAITGGSGITGMAVEAMNHAGVNNMLVTILND 180
 QY 181 DMSISPPVGAMNNYLTKVLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPGT 240
 Db 181 DMSISPPVGAMNNYLTKVLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPGT 240
 QY 241 LPEELGFNFGPDIHGDMVEMVSTLENLKDITGVPVFLHVTYKKGVAAPAKDPLAHGV 300
 Db 241 LPEELGFNFGPDIHGDMVEMVSTLENLKDITGVPVFLHVTYKKGVAAPAKDPLAHGV 300
 QY 301 PAFDPTKDFLPAKAPSPPHTYTEVFGRMLCDMAADERLLGITPAMREGSGLVEFSQKFP 360
 Db 301 PAFDPTKDFLPAKAPSPPHTYTEVFGRMLCDMAADERLLGITPAMREGSGLVEFSQKFP 360
 QY 361 NRYFDVAIAEQHAAVTLAAGACQCAKPVVAIYSTFLQRYDQILHDVALQNLDMFLALDR 420
 Db 361 NRYFDVAIAEQHAAVTLAAGACQCAKPVVAIYSTFLQRYDQILHDVALQNLDMFLALDR 420
 QY 421 AGLVGDPGPTHAGAFDYSYMRCLPNNLIMAPADNECRQMLTTGPOHGPASVRYPRGKG 480
 Db 421 AGLVGDPGPTHAGAFDYSYMRCLPNNLIMAPADNECRQMLTTGPOHGPASVRYPRGKG 480
 QY 481 PGAADPTLTALBIGAEVRHSGRIAILAMGSMVTPAVEAGKOLGATVNNRFPVKPDQ 540
 Db 481 PGAADPTLTALBIGAEVRHSGRIAILAMGSMVTPAVEAGKOLGATVNNRFPVKPDQ 540
 QY 541 ALVIELARTHDVPTVVEENVVIAAGAGSAINTEFLQAKVMPVCNIGLPDRFVEQSSREL 600
 Db 541 ALVIELARTHDVPTVVEENVVIAAGAGSAINTEFLQAKVMPVCNIGLPDRFVEQSSREL 600
 QY 601 LSLVGLDSKGLIATIEQFCA 620
 Db 601 LSLVGLDSKGLIATIEQFCA 620

RESULT 3 US-09-941-947a-6

; Sequence 6, Application US/09941947A
 ; Publication No. US2003000528A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Brzostowicz, Patricia C.
 ; APPLICANT: Cheng, Qiong
 ; APPLICANT: DiCosimo, Deana J.
 ; APPLICANT: Kofas, Matcheos
 ; APPLICANT: Miller, Edward S. Jr.
 ; APPLICANT: Odum, J. Martin
 ; APPLICANT: Picateggio, Steve
 ; APPLICANT: Rouviere, Pierre E.
 ; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
 ; FILE REFERENCE: CL1903 US NA
 ; CURRENT APPLICATION NUMBER: US/09/941,947A
 ; CURRENT FILING DATE: 2001-09-01
 ; PRIOR APPLICATION NUMBER: 60/229,907
 ; PRIOR FILING DATE: 2000-09-01
 ; PRIOR APPLICATION NUMBER: 60/229,858
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 60
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 6
 ; LENGTH: 620
 ; TYPE: PRT
 ; ORGANISM: Methylomonas 16a
 US-09-941-947a-6

Query Match 100.0%; Score 3216; DB 11; Length 620;
 Best Local Similarity 100.0%; Pred. No. 1,le-298;
 Matches 620; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLTIDYPLKNIHTPADIRALSKDQLQQLADEVGVLTHTVSISGGHFAAGLGTVELTV 60
 Db 1 MGLTIDYPLKNIHTPADIRALSKDQLQQLADEVGVLTHTVSISGGHFAAGLGTVELTV 60
 QY 61 ALHYVFNTPVDQLVWDVGHQAYPHKILTGKREMPITRTLGVSAPPADBSFYDAFVG 120

```

Qy 1 MKLITDYLKNIHTPADIRALSKDQCOQLADEVRGYLTHTVISGGHFAAGLGVETLV 60
Db 1 MKLITDYLKNIHTPADIRALSKDQCOQLADEVRGYLTHTVISGGHFAAGLGVETLV 60
Qy 61 ALATVFNTPVDQVMDVGHQAYPHKILTGKREKPTIRTLGGVSAFPARDESEYDAFGV 120
Db 61 ALATVFNTPVDQVMDVGHQAYPHKILTGKREKPTIRTLGGVSAFPARDESEYDAFGV 120
Qy 121 HSSSTISALAMALASOLRGEDKKVVAIIIGDSITGGAAVEAMHAGDVNNLLVTLNDN 180
Db 121 HSSSTISALAMALASOLRGEDKKVVAIIIGDSITGGAAVEAMHAGDVNNLLVTLNDN 180
Qy 181 DMSISPPVAGANNVLTLYSSKFPYSVVEESKKALAKPVSWEELARKTEEHVKGMIVPGT 240
Db 181 DMSISPPVAGANNVLTLYSSKFPYSVVEESKKALAKPVSWEELARKTEEHVKGMIVPGT 240
Qy 241 LPEELGFNYFPPIDGHVEMLVSTLENKDLTGPFPLHVTYTKKGKGYAPAEKDLAYHGV 300
Db 241 LPEELGFNYFPPIDGHVEMLVSTLENKDLTGPFPLHVTYTKKGKGYAPAEKDLAYHGV 300
Qy 301 PAPPTKDFLFRAPSPHPTYTEVGRMLCDMAADERLIGITPAMREGSGLVESQKFP 360
Db 301 PAPPTKDFLFRAPSPHPTYTEVGRMLCDMAADERLIGITPAMREGSGLVESQKFP 360
Qy 361 NRYFDVAIAEGHAYVTLAAGACQGAKEPVVAIYSTFLRGYDQLIHVALQNLMLFALDR 420
Db 361 NRYFDVAIAEGHAYVTLAAGACQGAKEPVVAIYSTFLRGYDQLIHVALQNLMLFALDR 420
Qy 421 AGLVGPPQPTNAGAPDVSVMCIENMLIMAPADNEHCQMLTTGFQHHGPASVYPRKGG 480
Db 421 AGLVGPPQPTNAGAPDVSVMCIENMLIMAPADNEHCQMLTTGFQHHGPASVYPRKGG 480
Qy 481 PGAAIDPTLFLAIGKAEVRHSGRIAILAMGSAVTPAVEAGKQAGATVVMRFVKPFDQ 540
Db 481 PGAAIDPTLFLAIGKAEVRHSGRIAILAMGSAVTPAVEAGKQAGATVVMRFVKPFDQ 540
Qy 541 ALVLELARTHDVFTVEBNVJAGAGSAINTFLOAQKYLMPVCNIGLDFRVEGGSREEL 600
Db 541 ALVLELARTHDVFTVEBNVJAGAGSAINTFLOAQKYLMPVCNIGLDFRVEGGSREEL 600
Qy 601 LSLVGLDSKGLATTEQFCA 620
Db 601 LSLVGLDSKGLATTEQFCA 620

RESULT 4
US-10-369-493-13998
; Sequence 13998, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13998
; LENGTH: 628
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-13998
Query Match 64.7%; Score 2080; DB 12; Length 628;
Best Local Similarity 63.2%; Pred. No. 76-190;
Matches 387; Conservative 101; Mismatches 120; Indels 4; Gaps 2;

```

```

Qy 8 PLIKNIHTPADIRALSKDQCOQLADEVRGYLTHTVISGGHFAAGLGVETLVALHAYFN 67
Db 16 PLIKNIHTPADIRALSKDQCOQLADEVRGYLTHTVISGGHFAAGLGVETLVALHAYFN 75
Qy 68 TPVDQVMDVGHQAYPHKILTGKREKPTIRTLGGVSAFPARDESEYDAFGVSHSSTIS 127
Db 76 TPVDQVMDVGHQAYPHKILTGKREKPTIRTLGGVSAFPARDESEYDAFGVSHSSTIS 135
Qy 128 AALGMAIALSOLRGEDKKVVAIIIGDSITGGAAVEAMHAGDVNNLLVTLNDNMSISPP 187
Db 136 AALGMAIALSOLRGEDKKVVAIIIGDSITGGAAVEAMHAGDVNNLLVTLNDNMSISPP 195
Qy 188 VGMANNVLTLYSSKFPYSVVEESKKALAKPVSWEELARKTEEHVKGMIVPGTLPEELGF 247
Db 196 VGMANNVLTLYSSKFPYSVVEESKKALAKPVSWEELARKTEEHVKGMIVPGTLPEELGF 255
Qy 248 NYGPIIDGHVEMLVSTLENKDLTGPFPLHVTYTKKGKGYAPAEKDLAYHGVPAEDPTK 307
Db 256 NYGPIIDGHVEMLVSTLENKDLTGPFPLHVTYTKKGKGYAPAEKDLAYHGVPAEDPTK 315
Qy 308 DFLFRAPSPHPTYTEVGRMLCDMAADERLIGITPAMREGSGLVESQKFPNRYFD 365
Db 316 --APPAAPKAGGPKTSGVGRMLCDMAADERLIGITPAMREGSGLVESQKFPNRYFD 373
Qy 366 VALAEGHAYVTLAAGACQGAKEPVVAIYSTFLRGYDQLIHVALQNLMLFALDRAGLV 425
Db 374 VALAEGHAYVTLAAGACQGAKEPVVAIYSTFLRGYDQLIHVALQNLMLFALDRAGLV 433
Qy 426 PDPTNAGAPDVSVMCIENMLIMAPADNEHCQMLTTGFQHHGPASVYPRKGGPRAI 485
Db 434 EDEPTNAGAPDVSVMCIENMLIMAPADNEHCQMLTTGFQHHGPASVYPRKGGPRAI 493
Qy 486 DPLFLAIGKAEVRHSGRIAILAMGSAVTPAVEAGKQAGATVVMRFVKPFDQALVLE 545
Db 494 EXLEPTEIGKGVRRQGSVALLVFGVMAELKXATLDAIVVMRFVKPFDQALVLE 553
Qy 546 LARTHDVFTVEBNVJAGAGSAINTFLOAQKYLMPVCNIGLDFRVEGGSREELSLVGL 605
Db 554 LARTHDVFTVEBNVJAGAGSAINTFLOAQKYLMPVCNIGLDFRVEGGSREELSLVGL 613
Qy 606 LDSKGLATTEQ 617
Db 614 LDSKGLATTEQ 625

RESULT 5
US-10-381-779-15
; Sequence 15, Application US/10381779
; Publication No. US2003021978A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jensen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002081
; CURRENT APPLICATION NUMBER: US/10/381,779
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatcSEO for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-381-779-15
Query Match 64.4%; Score 2069.5; DB 12; Length 627;
Best Local Similarity 63.4%; Pred. No. 7,1e-189;
Matches 388; Conservative 100; Mismatches 117; Indels 7; Gaps 2;

```

```

QY 8 PTLKNIHTPADIRALSKDQLQDLADVRGYLTHVTISGCHPAAGLGTVELTVALHYEN 67
DB 16 PLDDASSPAERIRLGRADLTLDLDEIKOYLTYTQYSGHAGLGVETLITLHYFD 75
QY 68 TPVDOLVMDVGHQAVPHKILTRKERMPITRTLGVSAPADESEYDAFGVGHSTIS 127
DB 76 TPDDLVMDVGHQAVPHKILTRERELMGTIRQKMLAEPRAESEYDTPGVGHSTIS 135
QY 128 AALGMAIASQLGEPKRWAIIGDSINGMAVEAMNAGVYANLVTILNDMSISP 187
DB 136 AALGMAIASQLGEPKRWAIIGDSINGMAVEAMNAGVYANLVTILNDMSISP 195
QY 188 VGAMNNYLYTKVLSKRYSSVREESKALAKMPSVMEARKTEEHYKGMIVPGLTPEELG 247
DB 196 VGLSNLYLAKLISSTYSSMRGSKVLSRLPGAMEIARTEHYAKMLVPGTLFEEIGM 255
QY 248 NTFGPIDGHDVEMLVSTLENKDLTGVPFLVYTKKGYAPAEKDPVLAHGVAPDPTK 307
DB 256 NTFGPIDGHDVEMLVSTLENKDLTGVPFLVYTKKGYAPAEKDPVLAHGVAPDPTK 310
QY 308 DFLPKAAP--SPHPTVTFVGRWLCMAAODERLLGITPAMRBSGLVEFSOKFENRYED 365
DB 311 LEAPGSAPEKKTGEPKTSYFGOMLCMAAODARLLGITPAMRBSGLVAFSERIPEKTFD 370
QY 366 VALAEQHAVTLAAGACQCAKPVVAIYSTLQRYGDIHVAALQNDMLPALDRAGLVG 425
DB 371 VALAEQHAVTLAAGACQCAKPVVAIYSTLQRYGDIHVAALQNDMLPALDRAGLVG 430
QY 426 PEGPHTAGAFDYSWRCIPNMLIMAPADNECROMLTGFGHGAASRYTRGKPGAAI 485
DB 431 EGGPHTAGAFDYSWRCIPNMLIMAPADNECROMLTGFGHGAASRYTRGKPGAAI 490
QY 486 DPTLTALBICKAEVRHHSRIATLAWGSMVTPAVEAKOLGATVVMRFPKPDQALVLE 545
DB 491 DPTLTALBICKAEVRHHSRIATLAWGSMVTPAVEAKOLGATVVMRFPKPDQALVLE 550
QY 546 LARTHDVFTVEENVLAGAGSAINTEFLQAKVIMPCNIGLDFRFEQGSREELSLVG 605
DB 551 LARTHDVFTVEENVLAGAGSAINTEFLQAKVIMPCNIGLDFRFEQGSREELSLVG 610
QY 606 LDSKGLIATTEQ 617
DB 611 LDAAGIEKAVRQ 622

```

RESULT 6
US-09-815-242-14083, Application US/09815242
Sequence 14083, Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27

```

PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 1410
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14083
LENGTH: 620
TYPE: PRF
ORGANISM: Salmonella typhi
US-09-815-242-14083

```

Query Match 62.6%; Score 2014.5; DB 9; Length 620;
Best local similarity 61.0%; Pred. No. 13e-183;
Matches 375; Conservative 106; Mismatches 131; Indels 3; Gaps 3;

```

QY 7 YPLKNIHTPADIRALSKDQLQDLADVRGYLTHVTISGCHPAAGLGTVELTVALHYF 66
DB 8 YPLTALVDSTQELRLPKESLPKLCBELRRYLLDSVRSRGHSPASGIGTVELTVALHYV 67
QY 67 NTFPVDOLVMDVGHQAVPHKILTRKERMPITRTLGVSAPADESEYDAFGVGHSTSI 126
DB 68 NTFPVDOLVMDVGHQAVPHKILTRKERMPITRTLGVSAPADESEYDAFGVGHSTSI 127
QY 127 SAALGMAIASQLGEPKRWAIIGDSINGMAVEAMNAGVYANLVTILNDMSISP 186
DB 128 SAALGMAIASQLGEPKRWAIIGDSINGMAVEAMNAGVYANLVTILNDMSISP 187
QY 187 PVGAMNNYLYTKVLSKRYSSVREESKALAKMPSVMEARKTEEHYKGMIVPGLTPEELG 246
DB 188 PVGAMNNYLYTKVLSKRYSSVREESKALAKMPSVMEARKTEEHYKGMIVPGLTPEELG 247
QY 247 FNYGSPIDGHDVEMLVSTLENKDLTGVPFLVYTKKGYAPAEKDPVLAHGVAPDPT 306
DB 248 FNYGSPIDGHDVEMLVSTLENKDLTGVPFLVYTKKGYAPAEKDPVLAHGVAPDPT 307
QY 307 KDFLPKAPSPHPTVTFVGRWLCMAAODERLLGITPAMRBSGLVEFSOKFENRYED 366
DB 308 KDFLPKAPSPHPTVTFVGRWLCMAAODERLLGITPAMRBSGLVEFSOKFENRYED 367
QY 367 AIAEQHAVTLAAGACQCAKPVVAIYSTLQRYGDIHVAALQNDMLPALDRAGLVG 426
DB 367 AIAEQHAVTLAAGACQCAKPVVAIYSTLQRYGDIHVAALQNDMLPALDRAGLVG 426
QY 427 DPTLTALBICKAEVRHHSRIATLAWGSMVTPAVEAKOLGATVVMRFPKPDQALVLE 545
DB 427 DPTLTALBICKAEVRHHSRIATLAWGSMVTPAVEAKOLGATVVMRFPKPDQALVLE 545
QY 486 DPTLTALBICKAEVRHHSRIATLAWGSMVTPAVEAKOLGATVVMRFPKPDQALVLE 545
DB 487 DPTLTALBICKAEVRHHSRIATLAWGSMVTPAVEAKOLGATVVMRFPKPDQALVLE 545
QY 546 LARTHDVFTVEENVLAGAGSAINTEFLQAKVIMPCNIGLDFRFEQGSREELSLVG 605
DB 546 LARTHDVFTVEENVLAGAGSAINTEFLQAKVIMPCNIGLDFRFEQGSREELSLVG 605
QY 606 LDSKGLIATTEQ 620
DB 606 LDAAGIEKAVRQ 620

```

RESULT 7
US-10-369-493-481
Sequence 481, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 481
 LENGTH: 619
 TYPE: PRT
 ORGANISM: Xenorhabdus nematophilus
 US-10-369-493-481

Query Match 62.3%; Score 2004.5; DB 12; Length 619;
 Best Local Similarity 61.8%; Pred. No. 1,2e-182;
 Matches 378; Conservative 102; Mismatches 129; Indels 3; Gaps 3;

QY 7 YPLLNHTPADIRALSSKQLOOLADEVRGYLTHVTSISGHPAAGLGTVELTVALHYF 66
 DB 8 YFTLALAEETPEELRLPKDTLPKLCDELROFLNVSRSRGHFAAGLGTVELTVALHYV 67
 QY 67 NTPVDQVWDVGHQAYPHKILITGRKEMPTRTLGVSAPPARDESEYDAFGVGHSTSI 126
 DB 68 KTFPDLVWDVGHQAYPHKILITGRKEMPTRTLGVSAPPARDESEYDAFGVGHSTSI 127
 QY 127 SAALGMAIASQLRGDKRMVAIIIGDSITGGAYEAMNAGDVANLLVILINDMSISP 186
 DB 128 SAGLGMAIAAKHEDKRGKRTVCVIGDGAITAGAFAMNAGDIDPDLVILINDMSISE 187
 QY 187 PYGAMNNYLTLYLSKFSVRSSESKKALAKMPSVWEIARKEEHNKMTIVPGLTPEELG 246
 DB 188 NGALNNHNLALHLSGLYTLTRSGKVFSPNLPKELIKTEELHKMVAVPGLTPEELG 247
 QY 247 FNYFGPIDGHVEMLVSTLENLKDITGPFVFLHVYTKKGAYAPAEKDPALAYGVAPDPT 306
 DB 248 FNYIGPVGDHVALYLTQTLKMRBELKGPQLHMTKKGAYAPAEKDPISWNAVPKEDPS 307
 QY 307 KQFLPKAASPHEPTTYEVFGMLCDMAADRELLGITPAMEGSGLVPSQKFPKRYFDV 366
 DB 308 TGSIPKSSDT-RFTFSKIFGEMLCSEAADKCLMAITPAMEGSGMVFPSREYPPQYFDV 366
 QY 367 ALAEQAVNTLAAGAACQAGAPVVAIYSTFLQRYDOLHDAVALONLDMFLALDRAGLVGP 426
 DB 367 ALAEQAVNTLAAGAACQAGAPVVAIYSTFLQRYDOLHDAVALONLDMFLALDRAGLVGP 426
 QY 427 DQPTAGAFDYSYMCIPNMLIMAPADENECROMLTGFGH-HGPAVSEYPRKCGGA 485
 DB 427 DQPTAGAFDYSYMCIPNMLIMAPADENECROMLTGFGH-HGPAVSEYPRKCGGA 486
 QY 486 DPTLTALLETIGKAEVYRHSGRIALLMGSMVTPAVENGKQLGATVYMMRFKPFDOALV 545
 DB 487 QP-LEILPFGKGVIRQGERIALINFGTLLPDLQSAEALNATVMMRFKPLDDELVLE 545
 QY 546 LARLTHDVFYVEENVYAGAGSAINYFLQAKVMEVCNIGLDPFVEGSGREELSLV 605
 DB 546 MASRDLVLTLENNALMGAGSGVNEMLMOTGNSVSNLGLPDPFIPQGTQALHSDLG 605
 QY 606 LQDSKGLIATIEQ 617
 DB 606 LQDSKGLIATIEQ 617

RESULT 8

US-10-369-493-21174
 Sequence 21174, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 21174
 LENGTH: 620
 TYPE: PRT
 ORGANISM: Xenorhabdus nematophilus
 US-10-369-493-21174

Query Match 62.3%; Score 2003.5; DB 12; Length 620;
 Best Local Similarity 61.1%; Pred. No. 1.5e-182;
 Matches 377; Conservative 102; Mismatches 131; Indels 7; Gaps 4;

QY 7 YPLLNHTPADIRALSSKQLOOLADEVRGYLTHVTSISGHPAAGLGTVELTVALHYF 66
 DB 8 YFTLALAEETPEELRLPKDTLPKLCDELROFLNVSRSRGHFAAGLGTVELTVALHYV 67
 QY 67 NTPVDQVWDVGHQAYPHKILITGRKEMPTRTLGVSAPPARDESEYDAFGVGHSTSI 126
 DB 68 KTFPDLVWDVGHQAYPHKILITGRKEMPTRTLGVSAPPARDESEYDAFGVGHSTSI 127
 QY 127 SAALGMAIASQLRGDKRMVAIIIGDSITGGAYEAMNAGDVANLLVILINDMSISP 186
 DB 128 SAGLGMAIAAKHEDKRGKRTVCVIGDGAITAGAFAMNAGDIDPDLVILINDMSISE 187
 QY 187 PYGAMNNYLTLYLSKFSVRSSESKKALAKMPSVWEIARKEEHNKMTIVPGLTPEELG 246
 DB 188 NGALNNHNLALHLSGLYTLTRSGKVFSPNLPKELIKTEELHKMVAVPGLTPEELG 247
 QY 247 FNYFGPIDGHVEMLVSTLENLKDITGPFVFLHVYTKKGAYAPAEKDPALAYGVAPDPT 306
 DB 248 FNYIGPVGDHVALYLTQTLKMRBELKGPQLHMTKKGAYAPAEKDPISWNAVPKEDPS 307
 QY 307 KQFLPKAASPHEPTTYEVFGMLCDMAADRELLGITPAMEGSGLVPSQKFPKRYFDV 366
 DB 308 TGSIPKSSDT-RFTFSKIFGEMLCSEAADKCLMAITPAMEGSGMVFPSREYPPQYFDV 366
 QY 367 ALAEQAVNTLAAGAACQAGAPVVAIYSTFLQRYDOLHDAVALONLDMFLALDRAGLVGP 426
 DB 367 ALAEQAVNTLAAGAACQAGAPVVAIYSTFLQRYDOLHDAVALONLDMFLALDRAGLVGP 426
 QY 427 DQPTAGAFDYSYMCIPNMLIMAPADENECROMLTGFGH-HGPAVSEYPRKCGGA 483
 DB 427 DQPTAGAFDYSYMCIPNMLIMAPADENECROMLTGFGH-HGPAVSEYPRKCGGA 484
 QY 484 AIDPTLTALLETIGKAEVYRHSGRIALLMGSMVTPAVENGKQLGATVYMMRFKPFDOALV 543
 DB 485 QP-LEILPFGKGVIRQGERIALINFGTLLPDLQSAEALNATVMMRFKPLDDELVLE 543
 QY 544 LARLTHDVFYVEENVYAGAGSAINYFLQAKVMEVCNIGLDPFVEGSGREELSLV 603
 DB 544 LEIASHDMVLTLENNALMGAGSGVNEMLMOTGNSVSNLGLPDPFIPQGTQALHSDLG 603
 QY 604 VGLDSKGLIATIEQ 620
 DB 604 VGLDSKGLIATIEQ 620

RESULT 9

US-09-815-242-10068
 Sequence 10068, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsem, Kari L.
 APPLICANT: Zvekind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.

```

APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/151,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10068
LENGTH: 620
TYPE: PRF
ORGANISM: Escherichia coli
US-09-815-242-10068

```

```

Query Match 62.2%; Score 2000.5; DB 9; Length 620;
Best Local Similarity 60.2%; Pred. No. 2.8e-182;
Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

```

```

7 YLLKNHTPADIRALSKDLOQLADVRGYLTHVTSISGCHFAAGLGTVELTVALHYF 66
8 YTLALVSTQGEHLRKPESLPKLCDELRFLDSVSSSGHFAAGLGTVELTVALHYF 67
67 NTPVDQLVMDVGHQAVPHKILTGKRMPTIRTLGGVSAFPADESEYDAFVGHSSTSI 126
68 NTPFDQLMDVGHQAVPHKILTGKRMPTIRTLGGVSAFPADESEYDAFVGHSSTSI 127
127 SALGMAIASQLEGDEKKNVAIIIGDSITGGMAVEANNAAGVANNLVTILNDMSISP 186
128 SAGIGAAVAABEGKKNRRVVCVIGDAITAGMAFEANNAAGDIRPDMVTILNDMSISP 187
187 PGAMANNYITKVLSSKFYSVRESKALAKMPVSWELAKTBEHVKMIVPGLTPEELG 246
188 NVGALNNHIAQLSGKLYSLREGKRVPSGVPIKELAKTBEHVKMIVPGLTPEELG 247
247 FNTFPGIDGHVEMLVSTLENKDLTGVPFLHVVTKKGVAPAEKDPPLAHVGPAPDPT 306
248 FNTFPGIDGHVEMLVSTLENKDLTGVPFLHVVTKKGVAPAEKDPPLAHVGPAPDPT 307
307 KDFLPKAPSPHPTTYEVFGRWLCDMAADDERLIGITPAMREGSGLVEPSOKFPNRYPDV 366
308 SCGLPSSS-GGLPSSISKIFGDMWCETAKDKKALITPAMREGSGLVEPSOKFPNRYPDV 366
367 AIAEOHVAITLAAGQACQAKPVVAIYSTFLQGYDQILHVAQLONTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGQACQAKPVVAIYSTFLQGYDQILHVAQLONTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGQACQAKPVVAIYSTFLQGYDQILHVAQLONTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGQACQAKPVVAIYSTFLQGYDQILHVAQLONTDMLFALDRAGLVGP 426
427 DGPTHAGAFDYVYMCIPNMLIMAPADENECROMLITTFQGH-CPASVYRPRKGGQAAI 485
427 DGPTHAGAFDYVYMCIPNMLIMAPADENECROMLITTFQGH-CPASVYRPRKGGQAAI 485
427 DGPTHAGAFDYVYMCIPNMLIMAPADENECROMLITTFQGH-CPASVYRPRKGGQAAI 485
427 DGPTHAGAFDYVYMCIPNMLIMAPADENECROMLITTFQGH-CPASVYRPRKGGQAAI 485
486 DPTLALBGAERHHSRIAILAMGSMVTPAVEAGKQAGATVNNRFPKPDQALVLE 545
486 DPTLALBGAERHHSRIAILAMGSMVTPAVEAGKQAGATVNNRFPKPDQALVLE 545
486 DPTLALBGAERHHSRIAILAMGSMVTPAVEAGKQAGATVNNRFPKPDQALVLE 545
486 DPTLALBGAERHHSRIAILAMGSMVTPAVEAGKQAGATVNNRFPKPDQALVLE 545
487 TP-LEKLPKIGKIVRKRGEKALINFGTLMPEAKVAESLNAITLVDMRFVXPDLEALILE 545
487 TP-LEKLPKIGKIVRKRGEKALINFGTLMPEAKVAESLNAITLVDMRFVXPDLEALILE 545
487 TP-LEKLPKIGKIVRKRGEKALINFGTLMPEAKVAESLNAITLVDMRFVXPDLEALILE 545
487 TP-LEKLPKIGKIVRKRGEKALINFGTLMPEAKVAESLNAITLVDMRFVXPDLEALILE 545
546 LAKTHDVYTVVEENYIAGGASAITFLQAKVILMPVCNIGLPDRFVQGSREELSLVIG 605
546 LAKTHDVYTVVEENYIAGGASAITFLQAKVILMPVCNIGLPDRFVQGSREELSLVIG 605
546 LAKTHDVYTVVEENYIAGGASAITFLQAKVILMPVCNIGLPDRFVQGSREELSLVIG 605
546 LAKTHDVYTVVEENYIAGGASAITFLQAKVILMPVCNIGLPDRFVQGSREELSLVIG 605
546 MAASHEALVYVEENAIMGAGSVNEVIMARHKKVPVLTINIGLPDFIPQGTQGEWRABIG 605
546 MAASHEALVYVEENAIMGAGSVNEVIMARHKKVPVLTINIGLPDFIPQGTQGEWRABIG 605
546 MAASHEALVYVEENAIMGAGSVNEVIMARHKKVPVLTINIGLPDFIPQGTQGEWRABIG 605
546 MAASHEALVYVEENAIMGAGSVNEVIMARHKKVPVLTINIGLPDFIPQGTQGEWRABIG 605
606 LDSKGLIATIEQFCA 620
606 LDSKGLIATIEQFCA 620

```

```

DB 606 LIDAGMEAKTAWLA 620
RESULT 10
US-10-381-779-31
Sequence 31, Application US/10381779
Publication No. US20030219798A1
GENERAL INFORMATION:
APPLICANT: Gokarn, Ravi R
APPLICANT: Jensen, Holly
APPLICANT: Zidwick, Mary Jo
TITLE OF INVENTION: Isoprenoid Production
FILE REFERENCE: 12304/002051
CURRENT APPLICATION NUMBER: US/10/381,779
PRIOR FILING DATE: 2003-03-28
PRIOR APPLICATION NUMBER: PCT/US01/30328
PRIOR FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: 60/236,580
PRIOR FILING DATE: 2000-09-29
NUMBER OF SEQ ID NOS: 190
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31
LENGTH: 620
TYPE: PRF
ORGANISM: Escherichia coli
US-10-381-779-31

```

```

Query Match 62.2%; Score 2000.5; DB 12; Length 620;
Best Local Similarity 60.2%; Pred. No. 2.8e-182;
Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

```

```

7 YLLKNHTPADIRALSKDLOQLADVRGYLTHVTSISGCHFAAGLGTVELTVALHYF 66
8 YTLALVSTQGEHLRKPESLPKLCDELRFLDSVSSSGHFAAGLGTVELTVALHYF 67
67 NTPVDQLVMDVGHQAVPHKILTGKRMPTIRTLGGVSAFPADESEYDAFVGHSSTSI 126
68 NTPFDQLMDVGHQAVPHKILTGKRMPTIRTLGGVSAFPADESEYDAFVGHSSTSI 127
127 SALGMAIASQLEGDEKKNVAIIIGDSITGGMAVEANNAAGVANNLVTILNDMSISP 186
128 SAGIGAAVAABEGKKNRRVVCVIGDAITAGMAFEANNAAGDIRPDMVTILNDMSISP 187
187 PGAMANNYITKVLSSKFYSVRESKALAKMPVSWELAKTBEHVKMIVPGLTPEELG 246
188 NVGALNNHIAQLSGKLYSLREGKRVPSGVPIKELAKTBEHVKMIVPGLTPEELG 247
247 FNTFPGIDGHVEMLVSTLENKDLTGVPFLHVVTKKGVAPAEKDPPLAHVGPAPDPT 306
248 FNTFPGIDGHVEMLVSTLENKDLTGVPFLHVVTKKGVAPAEKDPPLAHVGPAPDPT 307
307 KDFLPKAPSPHPTTYEVFGRWLCDMAADDERLIGITPAMREGSGLVEPSOKFPNRYPDV 366
308 SCGLPSSS-GGLPSSISKIFGDMWCETAKDKKALITPAMREGSGLVEPSOKFPNRYPDV 366
367 AIAEOHVAITLAAGQACQAKPVVAIYSTFLQGYDQILHVAQLONTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGQACQAKPVVAIYSTFLQGYDQILHVAQLONTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGQACQAKPVVAIYSTFLQGYDQILHVAQLONTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGQACQAKPVVAIYSTFLQGYDQILHVAQLONTDMLFALDRAGLVGP 426
427 DGPTHAGAFDYVYMCIPNMLIMAPADENECROMLITTFQGH-CPASVYRPRKGGQAAI 485
427 DGPTHAGAFDYVYMCIPNMLIMAPADENECROMLITTFQGH-CPASVYRPRKGGQAAI 485
427 DGPTHAGAFDYVYMCIPNMLIMAPADENECROMLITTFQGH-CPASVYRPRKGGQAAI 485
427 DGPTHAGAFDYVYMCIPNMLIMAPADENECROMLITTFQGH-CPASVYRPRKGGQAAI 485
486 DPTLALBGAERHHSRIAILAMGSMVTPAVEAGKQAGATVNNRFPKPDQALVLE 545
486 DPTLALBGAERHHSRIAILAMGSMVTPAVEAGKQAGATVNNRFPKPDQALVLE 545
486 DPTLALBGAERHHSRIAILAMGSMVTPAVEAGKQAGATVNNRFPKPDQALVLE 545
486 DPTLALBGAERHHSRIAILAMGSMVTPAVEAGKQAGATVNNRFPKPDQALVLE 545
487 TP-LEKLPKIGKIVRKRGEKALINFGTLMPEAKVAESLNAITLVDMRFVXPDLEALILE 545
487 TP-LEKLPKIGKIVRKRGEKALINFGTLMPEAKVAESLNAITLVDMRFVXPDLEALILE 545
487 TP-LEKLPKIGKIVRKRGEKALINFGTLMPEAKVAESLNAITLVDMRFVXPDLEALILE 545
487 TP-LEKLPKIGKIVRKRGEKALINFGTLMPEAKVAESLNAITLVDMRFVXPDLEALILE 545
546 LAKTHDVYTVVEENYIAGGASAITFLQAKVILMPVCNIGLPDRFVQGSREELSLVIG 605
546 LAKTHDVYTVVEENYIAGGASAITFLQAKVILMPVCNIGLPDRFVQGSREELSLVIG 605
546 LAKTHDVYTVVEENYIAGGASAITFLQAKVILMPVCNIGLPDRFVQGSREELSLVIG 605
546 LAKTHDVYTVVEENYIAGGASAITFLQAKVILMPVCNIGLPDRFVQGSREELSLVIG 605
546 MAASHEALVYVEENAIMGAGSVNEVIMARHKKVPVLTINIGLPDFIPQGTQGEWRABIG 605
546 MAASHEALVYVEENAIMGAGSVNEVIMARHKKVPVLTINIGLPDFIPQGTQGEWRABIG 605
546 MAASHEALVYVEENAIMGAGSVNEVIMARHKKVPVLTINIGLPDFIPQGTQGEWRABIG 605
546 MAASHEALVYVEENAIMGAGSVNEVIMARHKKVPVLTINIGLPDFIPQGTQGEWRABIG 605
606 LDSKGLIATIEQFCA 620
606 LDSKGLIATIEQFCA 620

```

Db 606 LIDAGMEAKIKXWLA 620

RESULT 11

US-10-369-493-731
 ; Sequence 731, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 731
 ; LENGTH: 620
 ; TYPE: PR1
 ; ORGANISM: Escherichia coli
 US-10-369-493-731

Query Match 62.2%; Score 2000.5; DB 12; Length 620;
 Best Local Similarity 60.2%; Pred. No. 2.8e-182;
 Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

7 YPLKNIHTPADIRALSKDLOQLADENVGYLTHVTSISGHPAAGLGTVELTVAHY 66
 8 YPLALVDSSTQELRLPKESLPCDELRLRYLDSRSRSGHPASLGLVELTVAHY 67
 67 NTPVDLVWDVGHQAPPHKILTRKERMPTIRTLGVSAPAPADESEYDAPGVGSS 126
 68 NTPVDLVWDVGHQAPPHKILTRKERMPTIRTLGVSAPAPADESEYDAPGVGSS 127
 127 SAALGKAIASQLRGDKKMAVAIGDSITGMAVEAMNAGVNNLVTINDNMSTSP 186
 128 SAGIGAVLAKEGKRRKRVCGVGDALTAGMAFEMNBAQDRPDMVLINDNMSTSE 187
 187 PVGAMNNVLTKYLSSKFFSVRESKALAKMPVWELARKTEHYKXIVGTLFEELG 246
 188 NVGALNNHLLQSLSGLYSLRGGKVSVPPIKELKREBHIKXVAVGTLFEELG 247
 247 FNVGSIIDHDVEMVSTLENKDLTGVPFLHVVTKGKYAPAEKDLAYHGPAPFPT 306
 248 FNVGSIIDHDVEMVSTLENKDLTGVPFLHVVTKGKYAPAEKDLAYHGPAPFPT 307
 307 KDFLPAASPHPTTYEVFGWMLCDMAADBRLLGITPAMREGSGLVESQKFPNRPY 366
 308 SGLTPKSS-CGLPSYKIRIGDMLCETRAKDNKMLATTPAMREGSGLVESQKFPN 366
 367 AIAEGAVTLAAGAACQAKPVVAIYSTFLRGYDOLIHVALQNLMLPALDRAGLVP 426
 427 DEPTVAGADVSVMRCIPMLIMAPADENECROMLTGFOH-GRAPSVYPPKSGGAI 485
 427 DQGTGAGADVSVMRCIPMLIMAPADENECROMLTGFOH-GRAPSVYPPKSGGAI 486
 486 DEPTVAGADVSVMRCIPMLIMAPADENECROMLTGFOH-GRAPSVYPPKSGGAI 545
 487 TP-LEKLPICKGIVKRRGKELINFGTLPBAKVAESIANIVDMRVPVKDLALILE 545
 546 LARTHVPTVVENYVAGAGSALNTFLQAKTLMVPCNIGPDRFVREGSBEELSLVG 605
 546 MAASHALVTVENYVAGAGSALNTFLQAKTLMVPCNIGPDRFVREGSBEELSLVG 605
 606 LIDAGMEAKIKXWLA 620

Db 606 LIDAGMEAKIKXWLA 620

RESULT 12

US-10-369-493-15911
 ; Sequence 15911, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 15911
 ; LENGTH: 623
 ; TYPE: PR1
 ; ORGANISM: Xanthomonas campestris
 US-10-369-493-15911

Query Match 62.0%; Score 1992.5; DB 12; Length 623;
 Best Local Similarity 61.4%; Pred. No. 1.7e-181;
 Matches 378; Conservative 91; Mismatches 142; Indels 5; Gaps 3;

4 TTDYELKNIHTPADIRALSKDLOQLADENVGYLTHVTSISGHPAAGLGTVELTVA 63
 5 STRPRLSRITQPPDLRFEEDLTAVADSLRAVLISVSGSGHPAAGLGTVELTVA 64
 64 YVNTVDLVWDVGHQAPPHKILTRKERMPTIRTLGVSAPAPADESEYDAPGVGSS 123
 65 YVNTVDLVWDVGHQAPPHKILTRKERMPTIRTLGVSAPAPADESEYDAPGVGSS 124
 124 TSISALGMAIASQLRGDKKMAVAIGDSITGMAVEAMNAGVNNLVTINDNMSTSP 181
 125 TSISALGMAIASQLRGDKKMAVAIGDSITGMAVEAMNAGVNNLVTINDNMSTSP 184
 182 MSISPPVAMNNVLTKYLSSKFFSVRESKALAKMPVWELARKTEHYKXIVGTLFE 239
 185 MSISPPVAMNNVLTKYLSSKFFSVRESKALAKMPVWELARKTEHYKXIVGTLFE 244
 240 TLFEELGNTYPTIDHDVEMVSTLENKDLTGVPFLHVVTKGKYAPAEKDLAYH 299
 245 TLFEELGNTYPTIDHDVEMVSTLENKDLTGVPFLHVVTKGKYAPAEKDLAYH 304
 300 VPAFPTDPLPKAPSPHPTTYEVFGWMLCDMAADBRLLGITPAMREGSGLVESQK 359
 305 VGPDPSPKGLVAKAG-AKKPTTYDVSVMCDMAADBRLLGITPAMREGSGLVESQ 363
 360 PNRYFDVAIAEGAVTLAAGAACQAKPVVAIYSTFLRGYDOLIHVALQNLMLPAL 419
 364 PNRYFDVAIAEGAVTLAAGAACQAKPVVAIYSTFLRGYDOLIHVALQNLMLPAL 423
 420 RAGVGPSPGTHAAGADVSVMRCIPMLIMAPADENECROMLTGFOH-GRAPSVY 479
 424 RAGVGPSPGTHAAGADVSVMRCIPMLIMAPADENECROMLTGFOH-GRAPSVY 483
 480 GPGALIDPTLALBEGKAVRHGSRILALMGSMVTPAVENAGKQAGATVNMRFVCFD 539
 484 GPGALIDPTLALBEGKAVRHGSRILALMGSMVTPAVENAGKQAGATVNMRFVCFD 543
 540 QALVLELARTHDVYVVENYVAGAGSALNTFLQAKTLMVPCNIGPDRFVREGSBE 599
 544 KAMLELAKCHEAVSVIEDNVVAGAGSALNTFLQAKTLMVPCNIGPDRFVREGSBE 603
 600 LIDAGMEAKIKXWLA 620

Db 566 KMLTLELACHEAFVIEDNVAGAGSVSEILNASTVLMPTLHGLDPSFGHNSRED 625
QY 600 LLSLVGLDSKGLIATI 615
Db 626 LLAEGIDQAGIRPAV 641

RESULT 15
US-10-369-493-7428

/ Sequence 7428, Application US/10369493
/ Publication No. US2003023675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ PRIOR FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 7428
/ LENGTH: 619
/ TYPE: PRT
/ ORGANISM: Burkholderia cepacia
US-10-369-493-7428

Query Match 61.9%; Score 1990.5; DB 12; Length 619;
Best Local Similarity 61.7%; Pred. No. 2.6e-181;
Matches 381; Conservative 89; Mismatches 141; Indels 7; Gaps 3;

QY 4 TTDYPLKXNHTPADRALSKDLOQLADVRGILTHTWVISGCHPAAGLGTVELTVALH 63
Db 3 TTMWDALKTIDDPALRRLLDRQLDELRAFLVDSVOTGHLSSNLGTVELTIALH 62
QY 64 YVNTFVDOGVMDVGHQAVPHKILTGKREMPRTITLGGVSAFPARDESEYDAFGVSHS 123
Db 63 YVDTFHDRIVMDVGHQAVPHKILTGKREMPRTITLGGVSAFPARDESEYDFTGTAHSS 122
QY 124 TSISALGMAIASQLEGEDEKQVAILIGDSITGSAVEANAG- DYNANLLVILLNDM 182
Db 123 TSISALGMAVASKLGKDRMGIAVIGDAMTAGAFAEMNNAGVEDVPLVILLNDM 182
QY 183 STSPYGAMNNTLTKLSTSFYSVAREBKKALAMPVWELAKTBEHYKGMIVPTLF 242
Db 183 STSPYGALNRHLARLSGPFYAARAVERVLRVAPMLDLARKLEHAKGMIVPATLF 242
QY 243 EELGFNYFGIDGHVEMVSTLENLKDITGPFVLHVETKKGKGYAPAEKDPPLAYHGA 302
Db 243 EELGFNYIGIDGHDLDSLITLQNKELRGPFVLHVETKKGKGYLAADPVLYHGRK 302
QY 303 FDETKDFLKAAPSPPHTTEVFGRMLCMAAODERLLGTTAMRRSGSLVESSQFPR 362
Db 303 FNAEGIKRPAATPS-KETTYQVGEWLCDAEELDARVIGITPAMREGSGMVEFEKFPDR 361
QY 363 YEDVALAEQHAATYLAAGACOGAKPVVALYSTFELRGYDQLHDVALQMLDMLFALDRAG 422
Db 362 YEDVGAIEQHAATYFAGGLAAGKPVVALYSTFELRAIDQLHDVALQMLPVVFAIDRAG 421
QY 423 LVGPDGPTAGAFDYSYMRCLPMLIMAPADENECHOMLTGFOHREGSAVEYPRGKPG 482
Db 422 LVGADGATAGAYDLFLRCIPMTVMYASDENECROMLYALQOPNPAVAYPRGAGTG 481
QY 483 AALDPTLTALFAGKAVRHH-----GSRITALLAMGSMTPAVEAGKOLGATVVMRFPYCP 537
Db 482 VAITIKOMTALPLJSGKIRRETSOPAKRIALILAFGTWVAPSLAAEQLDATVANKRFPYCP 541
QY 538 FQALVLELARTHDEVFTEVENVIAGAGSAINTPLAQCKVIMPVCNIGLPDRFVEQGR 597

Db 542 LDADIVRQLAETHDAIVTEEGCVAGAGSACVEALLASGTRPVLQGLPDRFIDHGP 601
QY 598 BELLSLVGLDSKGLIATI 615
Db 602 AMLAAGIDAVGITKSI 619

Search completed: January 29, 2004, 16:21:14
Job time : 44.3225 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:45:34 ; Search time 19.4586 Seconds
(without alignments)
1348.130 Million cell updates/sec

```
Title:      US-09-941-947A-6
Perfect score: 3216
Sequence:   1 MKLITDYPPLKNIHTPADIR.....LSIVGLDSKGLATIRQPCA 6200
```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : Issued Parents AA:*
1 : /cgm2_6/prodata/1/aa/5A_COMB.pep:*
2 : /cgm2_6/prodata/1/aa/5B_COMB.pep:*
3 : /cgm2_6/prodata/1/aa/6A_COMB.pep:*
4 : /cgm2_6/prodata/1/aa/6B_COMB.pep:*
5 : /cgm2_6/prodata/1/aa/PTCTs_COMB.pep:*
6 : /cgm2_6/prodata/1/aa/bacfilltest.pep:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2069.5	64.4	639	4	US-09-232-991A-26388	Sequence 26388, Appl1
2	1751	54.4	648	4	US-09-328-352-6210	Sequence 6210, Appl1
3	1485.5	46.2	720	4	US-09-857-556A-26	Sequence 26, Appl1
4	1428	44.4	594	4	US-09-857-556A-34	Sequence 34, Appl1
5	1420.5	44.2	708	4	US-09-857-556A-12	Sequence 12, Appl1
6	1418.5	44.1	724	3	US-09-146-221-6	Sequence 6, Appl1
7	1418.5	44.1	727	3	US-09-146-221-8	Sequence 8, Appl1
8	1411.5	43.9	659	4	US-09-626-569-2	Sequence 2, Appl1
9	1411.5	43.9	717	4	US-09-626-569-1	Sequence 1, Appl1
10	1411.5	43.9	824	4	US-09-626-569-2	Sequence 3, Appl1
11	1410.5	43.9	721	4	US-09-857-556A-10	Sequence 10, Appl1
12	1397.5	43.5	719	4	US-09-857-556A-33	Sequence 33, Appl1
13	1392	43.3	721	3	US-09-146-221-4	Sequence 4, Appl1
14	1148	35.7	644	4	US-09-198-452A-1135	Sequence 1135, Appl1
15	1086.5	33.8	736	4	US-08-311-731A-56	Sequence 56, Appl1
16	1086.5	33.8	758	4	US-08-311-731A-290	Sequence 290, Appl1
17	966	30.0	578	4	US-09-857-556A-28	Sequence 28, Appl1
18	917.5	28.5	458	4	US-09-857-556A-32	Sequence 32, Appl1
19	901.5	28.0	435	4	US-09-857-556A-20	Sequence 20, Appl1
20	893.5	27.8	597	4	US-09-107-532A-7053	Sequence 7053, Appl1
21	486	15.1	262	4	US-09-857-556A-18	Sequence 18, Appl1
22	424.5	13.2	211	4	US-09-857-556A-2	Sequence 2, Appl1
23	424	10.3	153	4	US-09-857-556A-22	Sequence 22, Appl1
24	331	10.3	343	4	US-09-328-352-5814	Sequence 5814, Appl1
25	280	8.7	121	4	US-09-857-556A-30	Sequence 30, Appl1
26	270.5	8.4	131	4	US-09-857-556A-4	Sequence 4, Appl1
27	265.5	8.3	540	3	US-09-011-074-4	Sequence 4, Appl1

ALIGNMENTS

28	240.5	7.5	675	4	US-09-134-001C-5219	Sequence 5219, Appl
29	229.5	7.1	692	4	US-09-199-452A-966	Sequence 966, Appl
30	225	7.0	124	4	US-09-185-556A-16	Sequence 16, Appl
31	213.5	6.6	655	3	US-09-382-106-2	Sequence 2, Appl 1
32	213.5	6.6	782	4	US-09-252-991A-19811	Sequence 19811, Appl
33	201.5	6.3	649	4	US-09-107-532A-4258	Sequence 4258, Appl
34	197.5	6.1	663	4	US-09-328-352-6441	Sequence 6441, Ap
35	182.5	5.7	655	4	US-09-298-724-2	Sequence 2, Appl 1
36	178.5	5.6	363	3	US-09-108-020-44	Sequence 44, Appl
37	176.5	5.5	337	4	US-09-134-001C-4640	Sequence 4640, Ap
38	170.5	5.3	337	4	US-09-134-001C-3207	Sequence 3207, Ap
39	164.5	5.1	702	6	5240838-15	Patent No. 5240838
40	155.5	5.0	366	3	US-09-108-020-46	Sequence 46, Appl
41	159	4.9	359	3	US-09-108-020-45	Sequence 45, Appl
42	156	4.8	685	4	US-09-252-991A-32382	Sequence 32382, Appl
43	153	4.8	375	3	US-09-108-020-54	Sequence 54, Appl
44	152	4.7	300	4	US-09-134-001C-3442	Sequence 3442, Ap
45	150.5	4.7	325	3	US-09-108-020-49	Sequence 49, Appl

RESULT 1
US-09-25

Sequence 26388, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 26388
 LENGTH: 639
 TYPE: PR1
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-26388

Query Match	64.4%	Score 2069.5	DB 4	Length 639
Best Local Similarity	63.4%	Pred. No. 4e-193		
Matches 388	Conservative 100	Mismatches 117	Indels 7	Gaps 2
QY	8	PILNKHTPPDIRALSKDQLQQLADYRGYLTHTTWSISGGHFAAGLGTVELTVALHYFN	67	
Db	28	PLIDRASSPHEIRLGRADLETTLDELRYLLTYVGQGGHFGAGLGVETLTALHYFD	87	
QY	68	TPVDQLVMDVGHQAYPKHLLTGERKMPITITLGGVSAFPARDESEYDAFGVHSSTIS	127	
Db	88	TPDRLVMDVGHQAYPKHLLTERELKLTGLQKNGLAAPRARSEYDTPFGVHSSTIS	147	
QY	128	AALGMALASQLRGDKKCAVILGGSGITGGHAYEAMNHAGVYANLVTINDNMSISDP	187	
Db	148	AALGMALAAARISQGERKSAVILGGALTAGDAFPAALNBAASVDMDVLTINDNMSISHN	207	
QY	188	VGAMNNLTITVLSSKFPSSVRAESBKKALAKMPVWEIARKEEYHKGWIVPQLEFEEIGF	247	
Db	208	VGGLSNVTLALISLRTYSSMRKESKYLSTRFGAMEIARKEEYFAKGLVYQTLFEEIGM	267	
QY	248	NIFQPIDGHDEVKLVSFTLENIKDTIGVFVLHVYTKKSGIYAPAKDPLAYHGPAPDPDK	307	
Db	268	NYIGPIDGHDLPTLVATLRNMRDMKGGQFHLVYTKKSGFAPALDPIGYAI-----TK	322	
QY	308	DELPKAAP--SPHPTTYEVFGKMLCDMAADDERLLGTPPARBSSGLVEFGKPNRYED	365	
Db	323	LEAPGSAFKKTGGKGVSSVFGQMLCDMAAQARLLGITPAKBSSDIYASERKPERFYD	382	

QY	366	VA:AEQAAATLAAGACACGAAKPEVVAIYSFELRGYDOLIHVVALQNDMLFALDRALG	425
	383	VA:AEQAAATLAAGACACGAAKPEVVAIYSFELRGYDOLIHVVAQYHDVLFALDRALG	442
Db	426	PDGPTTAGA:ADYSYMKCIPNMLIMADADENECROMLTTGPOHHPASVRYPRGKPGAAI	485
QY	443	EDGPTTAGSGFDISYLFNCIFPGMLVMTPSDEDELTKLLITGYLFDGPAVAVRYPRGGSPNHP	502
Db	486	DPFLTALAEIKGAVERHHGSRILAILMAGSWVTPAYAGKQLATAYVNNRFPVYKPGQALYLE	545
QY	503	DPDLO:PEVEICKGVYRRGGRVYALLPFGVQLAAKMYAASLDAITYVDNRFPVYKPLDHALVRE	562
Db	546	IAATHDVFTVEENVTIAGAGSAINITPLQAQVLMVQNICGLPDRFVYGQSGREBELLSLVG	605
QY	563	IAGSHELLVTIENNAVMGAGSAYGSEFLASBELIEPLQLGLPDTVEHAPDSHMLAECG	622
Db	606	LDKSGILATIEQ	617
QY	623	LDPAAGIEKAVRQ	634

```

RESULT 2
US-09-328-352-6210
: Sequence 6210, Application US/09328352
: Patent No. 6562958
: GENERAL INFORMATION:
: APPLICANT: Gary L. Breton et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
: TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC99-03PA
: CURRENT APPLICATION NUMBER: US/09/328,352
: CURRENT FILING DATE: 1999-06-04
: NUMBER OF SEQ ID NOS: 8252
: SEQ ID NO 6210
: LENGTH: 648
: TYPE: PR1
: ORGANISM: Acinetobacter baumannii
US-09-328-352-6210

```

Query March	54.43;	Score 1751;	DB 4;	Length 648;
Best Local Similarity	55.13;	Pred. No. 4.6e-167;		
Matches 342;	Conservative 94;	Mismatches 165;	Indels 20;	Gaps 6

[illegible]

Db 439 GPTGAGAYVAYMRITVPMVIMAPKDENECRQMLHTVVAANGPAKAYVPGAGVGVEIQK 498
 QY 488 TLTLALEIGAEV-----RHGSRDLIALMGSMVTPAYEAGKQ-----LGATVYNNRFVK 536
 Db 499 EKVYLEIGAEVIAEIKNSDEQITVLAFGSRVVVALEAAEQFAQKDVAVCVVNNRFVK 558
 QY 537 PFDQALVETLARTHTDVPVTEENVTAGAGSAINTPLQAKVMPVGNIGLIPDRPVEQGS 596
 Db 559 PDLQWITRIPLARTHTLPTVYESHALMGAGSAVNEPMHQQVATPIINLIGLSPFLHQAT 618
 QY 597 REELSLIVGLDSKGIATIEQ 617
 Db 619 HNGMLDQGLDKGIINSTER 639

```

RESULT 3
US-09-857-556A-26
: Sequence 26, Application US/09857556A
: Patent No. 6558915
:
: GENERAL INFORMATION:
: APPLICANT: Rebecca E. Cahoon
: APPLICANT: Sean J. Coughlan
: APPLICANT: Yong Tao
: APPLICANT: Zude Wang
: APPLICANT: Mark E. Williams
: TITLE OF INVENTION: Plant 1-Deoxy-Xyloose 5-phosphate Synthase
: PIR REFERENCE: BR1290
: CURRENT APPLICATION NUMBER: US/09/857,556A
: CURRENT FILING DATE: 2001-06-04
: PRIOR APPLICATION NUMBER: 60/110,779
: PRIOR FILING DATE: 1998-12-03
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: Microsoft Office 97
:
: SEQ ID NO 26
: LENGTH: 720
:
: TYPE: PRF
: ORGANISM: Oryza sativa
: US-09-857-556A-26

```

Query Match	46.2%	Score 1485.5;	DB 4;	length 720;
Best Local Similarity	49.2%;	Pred. No. 2.6e-140;		
Matches 314;	Conservative 102;	Mismatches 185;	Indels 37;	Gaps 13;

```

OY      8 PILKNIHTPADJFASLSKOOLADBEVGYJTHVYSGGGFAAGLSTVELTVAHVFN 67
Db      71 PLIDTVNYVTHKAKLSLKELQOLADBEHSVDI FHVYSKTGGHLSGLSGLVELTVAHVFN 130
OY      68 TPVDOLWMDVGHQAVPHKILTGKREKREFTITLLGVSAPFARDESEYDAFGVGHSTIS 127
Db      131 TPQDKILMDVGHQSIFPHILITGRBDKMPKMTQNTGLSGFTSRSESEYDPSGTGHSSITIS 190
OY      128 AALGVALISQLFGEDKRWALITGDSTGGWATYANQNHADVANKLVLIND--DMSI-- 184
Db      191 AALGVAAGDILGGRNNVAAVIGGAMTAGVAYEANNAGLIDSDMIVILINDKQVSLPT 250
OY      185 -----SPRYGMNNYTLTKVLSKFFYSVRESEKALLAKM--PSVWEIARKEEHWKMI- 236
Db      251 ATLDPAPRPAVGLSALSLSKQSSRLRELRKVAAGVTQKGGSVHEHLEAAVYDEARQMIS 310
OY      237 -VPGLTFEELGNYVEGPIIDGHVEMLVSTLEMLKD--LTGEVFLAVTTRKKGKGAAPAEKD 293
Db      311 GSGSTLFEELGLYITGPVDGHNIDILLILAEVKSSTTGGPEVLILHVTEKGRGIPVABERA 370
OY      294 PLAYHGVPAFDP--KDELPRAPSPHFTYTVVEGRMLCDMAADEBELLGITTPAREGSG 351
Db      371 ADKYGVAKFDPATGQF--KSPAKTLSTYNYFAZALILAEAGDQNRVAILHAAMGGGTG 427
OY      352 LVEFSQKPEPRKIFVDALAEQNAVTLAAGACQCGKPEVVALYSTLTQSGYQOULLHDVALQN 411
Db      428 LNVFLRPEPRNCFVDGIEQCHAVYFAAGLACBGKPKPCALYSSTLQSGYQOYVAVHDLOK 487
OY      412 LMDLFLADAGLVGSDGFTTHAGAFDYSYMRCTPNNLTLNAPADENB--RQRLTTFQGHNP 470

```

Db 488 LPRFAMDRAGLVGADGPTHGCAFDVTVNACLPNNVWVMAPSDEALICHVATTAALIDRRP 547
 QY 471 ASVRYPRGKGPAAIDPTL--TLEIGKAEVRHHSRIALLMGSVTVPAVEAGK----- 523
 Db 548 SCFRYRKGIGICVPLPNTKGVLEVGKGRVLLGGKRVALLGGSAVQYCLAAASIVENH 607
 QY 524 QLCATVYNNFVYKPEPDALVLELARTHDVFTVVEENVVAGGASAINTELQAKOXYL----- 579
 Db 608 GLKATVADARFCPLDQTLIRRLASSHEVLLTVBEGSI-GGFGSHVAQFMALDGLLDKLT 666
 QY 580 --MPVGNIGLDPDFVFGSGREELSLVGLDSKLIATI 615
 Db 667 KMRPLV--LPDSYIDHSGSPADQLAABAGLTPSHIAATV 701

RESULT 4

US-09-857-556A-34
 ; Sequence 34, Application US/09857556A
 ; Patent No. 6558915
 ; GENERAL INFORMATION:
 ; APPLICANT: Rebecca B. Cahoon
 ; APPLICANT: Sean J. Coughlan
 ; APPLICANT: Yong Tao
 ; APPLICANT: Zude Weng
 ; APPLICANT: Mark E. Williams
 ; TITLE OF INVENTION: Plant 1-Deoxy-Xylulose 5-Phosphate Synthase
 ; FILE REFERENCE: B81290
 ; CURRENT APPLICATION NUMBER: US/09/857,556A
 ; CURRENT FILING DATE: 2001-06-04
 ; PRIOR APPLICATION NUMBER: 60/110,779
 ; PRIOR FILING DATE: 1998-12-03
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 34
 ; LENGTH: 594
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 US-09-857-556A-34

Query Match 44.4%; Score 1428; DB 4; Length 594;
 Best Local Similarity 50.7%; Pred. No. 1,2e-134;
 Matches 298; Conservative 93; Mismatches 169; Indels 28; Gaps 11;

QY 16 PADIRALSDQQLADENVGTLTHTVSISGCHFAAGLSTVELTVALHYVNTPDQALV 75
 Db 3 PIMKRLSKKEQLADQLDELRSVDIFSVSKTGHGSSLSGVELTVALHYVNTPDQALV 62
 QY 76 DVGHQAYPHKILITGRKERMTITRTLGVSAPFAPADESYDAFGVSHSTISIAALGMATA 135
 Db 63 DVGHQAYPHKILITGRDKMTITRTLGVSAPFAPADESYDAFGVSHSTISIAALGMATA 122
 QY 136 SOLRGEDKQWVALIGDSITIGMAVYPAHHAGDVANLLVILNDN-DMSI-----SP 186
 Db 123 RLKGGKNNVAVIIGGAMTAGAYEAMNNAAGYLDSDMTIILNDNQVSLPATYLDGPAP 182
 QY 187 PVGAMNNYLTUKLSKEYSVREESKKALAKM-PSVWEIARKTEBEHYKMI--VPQTLE 243
 Db 183 PVGALLSSALSKQSSPLRLREVRKGVTKQIGSGVHELAAYVDEYARKMIGSSSTLE 242
 QY 244 ELGENTYFSDIGADVEMLVSTLENLD--LTGFVPLHVVTKKGGVAPAKDPLAYHGV 301
 Db 243 ELGLVYIGVDDGNIDDLITLREVSKTGTGVLIHVTEKRGVYAPABRAADKHHGA 302
 QY 302 ADPPT--KDFLPRAPSPHPTTEVGRMLCDMAADERLLGITPAMRGSGLVESQKF 359
 Db 303 KEDPATGKF--KSPAKTLSTYNNFAELIABEDNNVVAIHAAMGGSTELNYFLRRF 359
 QY 360 PRRYEDVAIAEGHVAVTLAAGACQAGAPVVAIYSTFLQGYDQLIHDAVLAOMLDMFLALD 419
 Db 360 PRRCGDVGAIEGHVAVTPAAGLACBGLKPKFCALYSSFLQGYDQVHDVLDQLPVRFAMD 419
 QY 420 BAGLVGPDGPTAGADVSYKRCIPNMLIMAPADENE-CRQMLTTFQGHGASVRYPRG 478

Db 420 BAGLVGADGPTHGCAFDVTVNACLPNNVWVMAPSDEALICHVATTAALIDRRSCFRYPRG 479
 QY 479 KGRGAIDPTL--TLEIGKAEVRHHSRIALLMGSVTVPAVEAGK-----QLCATVYN 531
 Db 480 NGIGPLPNTKGVLEVGKGRVLLGGKRVALLGGSAVQYCLAAASIVENHGLKATVAD 539
 QY 532 MRFVPEPDALVLELARTHDVFTVVEENVVAGGASAINTELQAKOXYL 579
 Db 540 ARFCPLDQTLIRRLASSHEVLLTVBEGSI-GGFGSHVAQFMALDGLD 586

RESULT 5

US-09-857-556A-12
 ; Sequence 12, Application US/09857556A
 ; Patent No. 6558915
 ; GENERAL INFORMATION:
 ; APPLICANT: Rebecca B. Cahoon
 ; APPLICANT: Sean J. Coughlan
 ; APPLICANT: Yong Tao
 ; APPLICANT: Zude Weng
 ; APPLICANT: Mark E. Williams
 ; TITLE OF INVENTION: Plant 1-Deoxy-Xylulose 5-Phosphate Synthase
 ; FILE REFERENCE: B81290
 ; CURRENT APPLICATION NUMBER: US/09/857,556A
 ; CURRENT FILING DATE: 2001-06-04
 ; PRIOR APPLICATION NUMBER: 60/110,779
 ; PRIOR FILING DATE: 1998-12-03
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 12
 ; LENGTH: 708
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 US-09-857-556A-12

Query Match 44.2%; Score 1420.5; DB 4; Length 708;
 Best Local Similarity 46.9%; Pred. No. 9.5e-134;
 Matches 300; Conservative 113; Mismatches 187; Indels 39; Gaps 14;

QY 8 PLKNHHPADIRALSDQQLADENVGTLTHTVSISGCHFAAGLSTVELTVALHYVNT 67
 Db 63 PLDQVNTVPIHMKRLSKKEQLADQLDELRSVDIFSVSKTGHGSSLSGVELTVALHYVNT 122
 QY 68 TPVDQVLDVGHQAYPHKILITGRKERMTITRTLGVSAPFAPADESYDAFGVSHSTIS 127
 Db 123 APDQILMDVGHQSPHKLITGRDKMTITRTLGVSAPFAPADESYDAFGVSHSTIS 182
 QY 128 AALGMAIASOLRGEDKQWVALIGDSITIGMAVYPAHHAGDVANLLVILNDN-DMSIS- 185
 Db 183 AGLGMAVGRDLKGRNNVAVIIGGAMTAGAYEAMNNAAGYLDSDMTIILNDNQVSLPT 242
 QY 186 -----PVGAMNNYLTUKLSKEYSVREESKKALAKMPS-VWEIARKTEBEHYKMI- 236
 Db 243 ATLDPPIPVGALLSSALSKRLQSNPLRLREVRKGVTKRIGSPHHELAAYVDEYARKMIS 302
 QY 237 -VPGTLFEEELGFNNFPGPIDGADVEMLVSTLENLD--LTGFVPLHVVTKKGGVAPAKD 293
 Db 303 GSGSSLFEELGLYIIGPDGNIDDLVAILNEVSKTGTGVLIHVTEKRGVYAPABRA 362
 QY 294 PLAYGVAPADPT--KDFLPRAPSPHPTTEVGRMLCDMAADERLLGITPAMRGSG 351
 Db 363 ADKHGVTKFDPDPGKQKSKATQSYTTY--FAELIABEADKXVVALHAAMGGTG 419
 QY 352 LVERSQKPRRYEDVAIAEGHVAVTLAAGACQAGAPVVAIYSTFLQGYDQLIHDAVLA 411
 Db 420 MLTHRRRPTTCFVGVIAIEGHVAVTPAAGLACBGLKPKFCALYSSFMORAYDDVHDVLD 479
 QY 412 LDMFLALDRAGLVGPDGPTAGADVSYKRCIPNMLIMAPADENE-CRQMLTTFQGHGAS 470
 Db 480 LPRFAMDRAGLVGADGPTHGCAFDVTVNACLPNNVWVMAPSDEALICHVATTAALIDRRP 539
 QY 471 ASVRYPRGKGPAAIDPTL--TLEIGKAEVRHHSRIALLMGSVTVPAVEAGK----- 523

Db 540 SCFRPGRNGIGVQJL-PTGNKGTPLEIGKRIIEGERVALLGYSAGVONCLAAASLVBC 598

Qy 524 -QUGAVNMRFRKPPDQALVLELARTHVYFVVEENVVAGGASALNTPLOAKTL--- 579

Db 599 HGLRLVADRFCKPLDRSLIRSLASHEVLIVVEGSI--GGGSHVAFMALDGLDCK 657

Qy 580 ---MPVCNIGLPDRFVEGGRRELLSLVGLDSKGLIATI 615

Db 658 LKMRPIV---LPDRYIDHGSPPADQSLAGTSPSHIATV 693

RESULT 6

US-09-146-221-6

Sequence 6, Application US/09146221

Patent No. 6190895

GENERAL INFORMATION:

APPLICANT: Croteau, Rodney B

APPLICANT: Lange, Bernd M

APPLICANT: Wildung, Mark R

APPLICANT: McCaskill, David G

TITLE OF INVENTION: Nucleic and Amino Acid Sequences Relating to a No. 6190895el

FILE REFERENCE: No. 6190895el transketolase from peppermint

CURRENT APPLICATION NUMBER: US/09/146,221

CURRENT FILING DATE: 1998-09-01

EARLIER APPLICATION NUMBER: 60/056,033

EARLIER FILING DATE: 1997-09-02

NUMBER OF SEQ ID NOS: 12

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 6

LENGTH: 724

TYPE: PRT

ORGANISM: Mentha piperita

US-09-146-221-6

Query Match 44.1%; Score 1418.5; DB 3; Length 724;

Best Local Similarity 46.9%; Pred. No. 1.6e-133;

Matches 298; Conservative 116; Mismatches 189; Indels 33; Gaps 13;

Qy 8 PLKNIHTPADIRALSKDQLOOLADEVRGYLTHVTSISGHPAAGLIVELVLAHYVN 67

Db 77 PILDITVYPMNKNLSVEELANLDELREIVTVSKTGHLSSLSGSELVLAHYVN 136

Qy 68 TPVDQVMDVGHQAYPHKILTGKREMPITITLGVSAFPADSEVDARVGHSSSTIS 127

Db 137 TPDKRIWVGHQAYPHKILTGKRAEMHTIRQTGLAGFPKDESAHDAFGAGHSSTIS 196

Qy 128 AALGMAIASOLRGEDKQVAILIGDSITGMAVEAMNHAGVNAVNLVILNDN-DMSI-- 184

Db 197 AGLGMAVARDLQKNNHVISYIGDGMATGQAYEALNNGFLDSNLIIVANDNQVSLPT 256

Qy 185 -----SPVGMANNVLTIVLSSKFSYSSVRESKALAMPDS-VWEIARKTEBHVKMI- 236

Db 257 ATVDGPAPVPVALSKALITLQASRKFRQLREAAKSMTKOMGAPAEHIASKLTQYVKGMG 316

Qy 237 VPG-TLFEELGFNYPFGIDGHVEMVSTLENKDL--TGPFVLAHYVTKGKGVAPARKD 293

Db 317 KEGASLFEELGTYIGFVDGHNVEDLVYIFKAVKEMPAEPVLIHITIKGKGPAPARIA 376

Qy 294 PLAYHGVAPFPTDPLPKAPSPHETTYEVFGRLCDMAADERLLGITPAMEBSGLV 353

Db 378 ADMKGVVAFD-AKTGKQMKTKNKTKSYTFQFASLVAEAEHDKIIVAHAMGGGTGLN 436

Qy 354 EFSQKFRMYRVDVIAAEOHVTLLAAGACGAKPVVAIYSTFLORGVDLIHVDALNLD 413

Db 436 IFQKQPPDRCDVGLAEOHVTLLAAGABGKLPCLATYSSFLRGVDVYHVDLQKLP 496

Qy 414 MLZALDRAGLVGPDPGTHAGAFDYSYMKCIPIMLIMAPADENECQML-TTGFQHHGPAS 472

Db 496 VRFMDRAGVVGADGPTHGAFDPTTMACTPNNVVAAPSDAEALNMATATAIIDRSC 555

Qy 473 VRYRGGKGAALDPTL--TALFEGKAEVHHSKRIALLANGSMYTPAVAGKOL----- 525

Db 556 VRYRNGNGIGVALLPSNNKGTPLBIGKRIILKBSKVAILIGSTTVQNCMAANLLEQHI 615

Qy 526 GATVNNMFVFRPDQALVLELARTHVYFVVEENVVAGGASALNTPLOAKTL----- 579

Db 616 SVTVADAFCKPLDRLDKLVQHEIRVLIYVEBSGI--GGFAHISHFLANGLDGNLKW 674

Qy 580 MPVCNIGLPDRFVEGGRRELLSLVGLDSKGLIATI 615

Db 675 RPMV---LPDRYIDHGSPPADQSLAGTSPSHIATV 707

RESULT 7

US-09-146-221-8

Sequence 8, Application US/09146221

Patent No. 6190895

GENERAL INFORMATION:

APPLICANT: Croteau, Rodney B

APPLICANT: Lange, Bernd M

APPLICANT: Wildung, Mark R

APPLICANT: McCaskill, David G

TITLE OF INVENTION: Nucleic and Amino Acid Sequences Relating to a No. 6190895el

FILE REFERENCE: No. 6190895el transketolase from peppermint

CURRENT APPLICATION NUMBER: US/09/146,221

CURRENT FILING DATE: 1998-09-01

EARLIER APPLICATION NUMBER: 60/056,033

EARLIER FILING DATE: 1997-09-02

NUMBER OF SEQ ID NOS: 12

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 8

LENGTH: 727

TYPE: PRT

ORGANISM: Mentha piperita

US-09-146-221-8

Query Match 44.1%; Score 1418.5; DB 3; Length 727;

Best Local Similarity 46.8%; Pred. No. 1.6e-133;

Matches 298; Conservative 116; Mismatches 190; Indels 33; Gaps 13;

Qy 8 PLKNIHTPADIRALSKDQLOOLADEVRGYLTHVTSISGHPAAGLIVELVLAHYVN 67

Db 78 PILDITVYPMNKNLSVEELANLDELREIVTVSKTGHLSSLSGSELVLAHYVN 137

Qy 68 TPVDQVMDVGHQAYPHKILTGKREMPITITLGVSAFPADSEVDARVGHSSSTIS 127

Db 138 TPDKRIWVGHQAYPHKILTGKRAEMHTIRQTGLAGFPKDESAHDAFGAGHSSTIS 197

Qy 128 AALGMAIASOLRGEDKQVAILIGDSITGMAVEAMNHAGVNAVNLVILNDN-DMSI-- 184

Db 198 AGLGMAVARDLQKNNHVISYIGDGMATGQAYEALNNGFLDSNLIIVANDNQVSLPT 257

Qy 185 -----SPVGMANNVLTIVLSSKFSYSSVRESKALAMPDS-VWEIARKTEBHVKMI- 236

Db 258 ATVDGPAPVPVALSKALITLQASRKFRQLREAAKSMTKOMGAPAEHIASKLTQYVKGMG 317

Qy 237 VPG-TLFEELGFNYPFGIDGHVEMVSTLENKDL--TGPFVLAHYVTKGKGVAPARKD 293

Db 318 KEGASLFEELGTYIGFVDGHNVEDLVYIFKAVKEMPAEPVLIHITIKGKGPAPARIA 377

Qy 294 PLAYHGVAPFPTDPLPKAPSPHETTYEVFGRLCDMAADERLLGITPAMEBSGLV 353

Db 378 ADMKGVVAFD-AKTGKQMKTKNKTKSYTFQFASLVAEAEHDKIIVAHAMGGGTGLN 436

Qy 354 EFSQKFRMYRVDVIAAEOHVTLLAAGACGAKPVVAIYSTFLORGVDLIHVDALNLD 413

Db 437 IFQKQPPDRCDVGLAEOHVTLLAAGABGKLPCLATYSSFLRGVDVYHVDLQKLP 496

Qy 414 MLZALDRAGLVGPDPGTHAGAFDYSYMKCIPIMLIMAPADENECQML-TTGFQHHGPAS 472

Db 497 VRFMDRAGVVGADGPTHGAFDPTTMACTPNNVVAAPSDAEALNMATATAIIDRSC 556

Qy 473 VRYRGGKGAALDPTL--TALFEGKAEVHHSKRIALLANGSMYTPAVAGKOL----- 525

Db 557 VRRPGRNGIGVALPNNKGTPLRIGKRIKESGKVALIFGTIVONCAAAATLEQHI 616
 Qy 526 GATVVMRRFPKPPDQALVLEARTHDVPTVEENVIA-GGASAIITFLQAKVL----- 579
 Db 617 STVADARCKPDPDLIKLVQHEVLTVEGSGIGIGSGSHISHFSLINGLDNLIK 676
 Qy 580 -MPVCNIGLPDRFVQSGREELSLVGLDSKGLIAT 615
 Db 677 KRPVW---LPDRYIDHQAQSDQIEEAGLSPLHAGIV 710

RESULT 8
 US-09-626-589-2
 ; Sequence 2, Application US/09626589
 ; Patent No. 6326164
 ; GENERAL INFORMATION:
 ; APPLICANT: Rice, John
 ; APPLICANT: Klotz, Andreas
 ; APPLICANT: Crawford, John
 ; APPLICANT: Lanning, Beth
 ; APPLICANT: Stewart, Sandy
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
 ; IDENTIFICATION OF MODULATORS OF DEOXYXYLUOSE 5-PHOSPHATE
 ; TITLE OF INVENTION: SYNTHASE ACTIVITY
 ; FILE REFERENCE: 2037 US
 ; CURRENT APPLICATION NUMBER: US/09/626,589
 ; CURRENT FILING DATE: 2000-07-27
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 659
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-626-589-2

Query Match 43.9%; Score 1411.5; DB 4; Length 659;
 Best Local Similarity 46.6%; Pred. No. 6.8e-133;
 Matches 297; Conservative 115; Mismatches 188; Indels 37; Gaps 13;

Qy 8 PLKNIHTPADIRALSKDQLQDLADVRGYLTHTVVISGSHFPAAGLGTVELTVALHYFN 67
 Db 17 PLDTINVTIMKNSVRELKQSLDELRSDDVFNVSKTGHLGSSLGVELTVALHYFN 76
 Qy 68 TPVDQLVMDVGHQAVPHKILTGKREKMPITRTLGVSAPPADESEYDAFGVGSSTIS 127
 Db 77 TPQDKILMDVGHQSPHKLITGRKGMPTMRTGTNGISGPTKGESEHDCFTGHSSTIS 136
 Qy 128 AALGMALASQLRGDKKQVATIGDSITGMAVEAMNHAGDVANLVTINDN-DMSI-- 184
 Db 137 AGLGMAVRGDLKGRNNVAVIGDAMTGAQAYEAMNNAGTLDSDVITLNDKQVSLPT 196
 Qy 185 -----SPVGAANNVLTVLSSKRYSSVREBSKKALAMPS-VWELARKTEEHVKMI- 236
 Db 197 ATLDGSPFVGLASSLSRLQSNPALREIREVAKGKTQKQIGPMHQLAAKVEYARGMIS 256
 Qy 237 -VPGLTPEELGNTYGPIDGDMVSTLENLKD--LTGPFVLTAVYTKKGTAPAEKD 293
 Db 257 GTGSSLPFELGLYIYIPVDGHNIDDLVALIKVSTRTTGPVLHVVTEKRGYPYARA 316
 Qy 294 PLAYGVAPADPT--KDFLPKAPSPHPTYTEVFGWLCMAAODERLLGITPMARESSG 351
 Db 317 DDKHGVVKTDPARGRP---KTNKQOSTYTYFAELVAMEVDKQVVAIHAMGGGTG 373
 Qy 352 LVBSQKFPNRYFPVVAIAEQAHTLAAGACQAGAPVAIYSTLQNGYDOLIHVALQN 411
 Db 374 LNLQRRFPFPCFVGIABQHAHTFAAGLACBGLKPCACAYSSFMORAYDQVVDVLDK 433
 Qy 412 LDMFALDRAGLVGDPGPTAGADVSVMRCIPNMLTMAPDEKSCOMLTGFC-HGCP 470
 Db 434 LPVAFADRAGLVGADGPTGCAFDVTFAACIPMNIWAPSDRALFKVATVAIADRP 493
 Qy 471 ASVRYPRGKRGAAIDP--TLTALIEGKAEVRHHSRIAILAMGSMVTPAVEAG-----K 523

Db 494 SCFRYPGRNGIGVALPNNKGVPIRIGKRIKESGERVALLGYGSAVQSCGAVALLEER 553
 Qy 524 QGATVVMRRFPKPPDQALVLEARTHDVPTVEENVIA-GGASAIITFLQAKVL----- 579
 Db 554 GLMTVAARCKPDPDLIRSLAKSHEVLTVEGSGI-GGGSHVVOFLADGLDGLK 612
 Qy 580 -MPVCNIGLPDRFVQSGREELSLVGLDSKGLIAT 614
 Db 613 KRPVW---LPDRYIDHQAQSDQIEEAGLSPLHAGIV 710

RESULT 9
 US-09-626-589-1
 ; Sequence 1, Application US/09626589
 ; Patent No. 6326164
 ; GENERAL INFORMATION:
 ; APPLICANT: Rice, John
 ; APPLICANT: Klotz, Andreas
 ; APPLICANT: Crawford, John
 ; APPLICANT: Lanning, Beth
 ; APPLICANT: Stewart, Sandy
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
 ; IDENTIFICATION OF MODULATORS OF DEOXYXYLUOSE 5-PHOSPHATE
 ; TITLE OF INVENTION: SYNTHASE ACTIVITY
 ; FILE REFERENCE: 2037 US
 ; CURRENT APPLICATION NUMBER: US/09/626,589
 ; CURRENT FILING DATE: 2000-07-27
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 717
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-626-589-1

Query Match 43.9%; Score 1411.5; DB 4; Length 717;
 Best Local Similarity 46.6%; Pred. No. 7.8e-133;
 Matches 297; Conservative 115; Mismatches 188; Indels 37; Gaps 13;

Qy 8 PLKNIHTPADIRALSKDQLQDLADVRGYLTHTVVISGSHFPAAGLGTVELTVALHYFN 67
 Db 75 PLDTINVTIMKNSVRELKQSLDELRSDDVFNVSKTGHLGSSLGVELTVALHYFN 134
 Qy 68 TPVDQLVMDVGHQAVPHKILTGKREKMPITRTLGVSAPPADESEYDAFGVGSSTIS 127
 Db 135 TPQDKILMDVGHQSPHKLITGRKGMPTMRTGTNGISGPTKGESEHDCFTGHSSTIS 194
 Qy 128 AALGMALASQLRGDKKQVATIGDSITGMAVEAMNHAGDVANLVTINDN-DMSI-- 184
 Db 195 AGLGMAVRGDLKGRNNVAVIGDAMTGAQAYEAMNNAGTLDSDVITLNDKQVSLPT 254
 Qy 185 -----SPVGAANNVLTVLSSKRYSSVREBSKKALAMPS-VWELARKTEEHVKMI- 236
 Db 255 ATLDGSPFVGLASSLSRLQSNPALREIREVAKGKTQKQIGPMHQLAAKVEYARGMIS 314
 Qy 237 -VPGLTPEELGNTYGPIDGDMVSTLENLKD--LTGPFVLTAVYTKKGTAPAEKD 293
 Db 315 GTGSSLPFELGLYIYIPVDGHNIDDLVALIKVSTRTTGPVLHVVTEKRGYPYARA 374
 Qy 294 PLAYGVAPADPT--KDFLPKAPSPHPTYTEVFGWLCMAAODERLLGITPMARESSG 351
 Db 375 DDKHGVVKTDPARGRP---KTNKQOSTYTYFAELVAMEVDKQVVAIHAMGGGTG 431
 Qy 352 LVBSQKFPNRYFPVVAIAEQAHTLAAGACQAGAPVAIYSTLQNGYDOLIHVALQN 411
 Db 432 LNLQRRFPFPCFVGIABQHAHTFAAGLACBGLKPCACAYSSFMORAYDQVVDVLDK 491
 Qy 412 LDMFALDRAGLVGDPGPTAGADVSVMRCIPNMLTMAPDEKSCOMLTGFC-HGCP 470
 Db 492 LPVAFADRAGLVGADGPTGCAFDVTFAACIPMNIWAPSDRALFKVATVAIADRP 551
 Qy 471 ASVRYPRGKRGAAIDP--TLTALIEGKAEVRHHSRIAILAMGSMVTPAVEAG-----K 523

Db 552 SCRRYPRNGNIGVALPFGKGVPIEIGKRIKKEGRVALLGSAVSGCLGAAMLEER 611
 Qy 524 QLGATVVMRFVPEFPOALVLELARTHDVPEVVEENVJAGAGSAINTELOAKVY---- 579
 Db 612 GLNVTADARFCCKPLRALIRLSAKSHEVLITVEBGSIGFGSHVVOFLALDGLDGL 670
 Qy 580 --MPVCNIGLPDRFVBOGSRBELSLVGLDSKGIAT 614
 Db 671 KMRPMV--LPDRYIDHGAPADQLAERGLMPSHIAT 704

RESULT 10
 US-09-626-589-3
 / Sequence 3, Application US/09626589
 / Patent No. 6326164
 / GENERAL INFORMATION:
 / APPLICANT: Rice, John
 / APPLICANT: Klotz, Andreas
 / APPLICANT: Crawford, John
 / APPLICANT: Lanning, Beth
 / APPLICANT: Stewart, Sandy
 / TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
 / TITLE OF INVENTION: IDENTIFICATION OF MODULATORS OF DEOXYXYLOSE 5-PHOSPHATE
 / FILE REFERENCE: 2037 US
 / CURRENT APPLICATION NUMBER: US/09/626,589
 / CURRENT FILING DATE: 2000-07-27
 / NUMBER OF SEQ ID NOS: 6
 / SOFTWARE: FASTSEQ for Windows Version 4.0
 / SEQ ID NO 3
 / LENGTH: 824
 / TYPE: PRF
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: Residues 1-165 are from the thioredoxin sequence
 / OTHER INFORMATION: found in the vector pBT32 (supplied by No. 6326164agen.
 / OTHER INFORMATION: Residues 166-824 represent the DXPS sequence from
 / OTHER INFORMATION: Arabidopsis shown in SEQ ID NO:2.
 US-09-626-589-3

Query Match 43.9%; Score 1411.5; DB 4; Length 824;
 Best Local Similarity 46.6%; Pred. No. 9.9e-133;
 Matches 297; Conservative 115; Mismatches 188; Indels 37; Gaps 13;

Qy 8 PLKNHITPADIRALSKQLOQLADEVRGILTHTVTSISGHHFAAGLCTVELTVALHYFN 67
 Db 182 PLADTINVPFHKNLSVVELKQLSDELRSDFVIFHVSRTGHLSSLGVELTVALHYFN 241
 Qy 68 TPVDOLVWDVGHQAVPHKILITGRKERMPITRTIGVSAFPARDESEYDAFGVGHSSSTIS 127
 Db 242 TPQDKIIMDVGHQSYFHKILITGRKRMPTMOTNLSGFTKRGSEHDCFGTGHSSSTIS 301
 Qy 128 AALGMAIASQLRSGEDKKMAVITIGDSITGMAVEAMNAGDVANANLVTILNDN-DMSI- 184
 Db 302 AGLGMAVAGDLKGNKNVAVITIGDAMTAGQAVEAMNAGVLDSDMVIILNDNKQVSLPT 361
 Qy 185 -----SPVGMANNVITKLSKFFYSVREESKXALAKMPS-VNELARKTEBHVKGMI- 236
 Db 362 ATLDPSPVPGALSSALSRLOSNPLRREIRVAKAKMTQKGGPHEQLAAKVDETARKMTIS 421
 Qy 237 -VPGTLFEELGFNYFGFIDGHEMLVSTLENKD-LTGPVFLAVVTKKKGAYAPAEKD 293
 Db 422 GTGSSLFEELGLVYIGVDSHNIDLVSLINEVKSRTTGPVLHVTEKRGVPAVERA 481
 Qy 294 PLAYHGVPADEPT--KDFLPKAPSPHPTTYEVFGKRLCMAADBERLGLITPMRBSGS 351
 Db 482 DDKHGVKVPDPTGRQF--KTTNKTSYTYFEBALVABAEVDKDVALLHAAMGGSTG 538
 Qy 352 LVEFSQKFPNRYPDVAIAEQHNAVTLAAGQACQAGAPVVAISTFLQSGYDOLIHDAVON 411
 Db 539 LNLFPGRFPTRCFDVGIAEQHNAVTFPAAGLAGEGLPFCALISSFMGRAYDQVVDVLDQK 598
 Qy 412 LKMLPALDRAGLVGDPGPTHAGAFDYSYMRCLPNMLIMAPADENECROMLTTGFO-HHGP 470

Db 599 LPVFPAMDRAGLVADGPTHGCAFVTFMACLPNMTVWAPSDADEFMVATAVAIDRRP 658
 Qy 471 ASVYPRKKGGAIDP--TITALEIGABVRHHGSRITALLAKMSMTPAVAG-----K 523
 Db 659 SCFRYPRNGIGVALPFGKGVPIEIGKRIKKEGRVALLGSAVSGCLGAAMLEER 718
 Qy 524 QLGATVVMRFVPEFPOALVLELARTHDVPEVVEENVJAGAGSAINTELOAKVY---- 579
 Db 719 GLNVTADARFCCKPLRALIRLSAKSHEVLITVEBGSIGFGSHVVOFLALDGLDGL 777
 Qy 580 --MPVCNIGLPDRFVBOGSRBELSLVGLDSKGIAT 614
 Db 778 KMRPMV--LPDRYIDHGAPADQLAERGLMPSHIAT 811

RESULT 11
 US-09-857-556A-10
 / Sequence 10, Application US/09857556A
 / Patent No. 6558915
 / GENERAL INFORMATION:
 / APPLICANT: Rebecca E. Cahoon
 / APPLICANT: Sean J. Coughlan
 / APPLICANT: Yong Tao
 / APPLICANT: Zude Keng
 / APPLICANT: Mark E. Williams
 / TITLE OF INVENTION: Plant 1-Deoxy-Xylose 5-Phosphate Synthase
 / FILE REFERENCE: BB1290
 / CURRENT APPLICATION NUMBER: US/09/857,556A
 / CURRENT FILING DATE: 2001-06-04
 / PRIOR APPLICATION NUMBER: 60/110,779
 / PRIOR FILING DATE: 1998-12-03
 / NUMBER OF SEQ ID NOS: 34
 / SOFTWARE: Microsoft Office 97
 / SEQ ID NO 10
 / LENGTH: 721
 / TYPE: PRF
 / ORGANISM: Glycine max
 US-09-857-556A-10

Query Match 43.9%; Score 1410.5; DB 4; Length 721;
 Best Local Similarity 47.0%; Pred. No. 1e-132;
 Matches 302; Conservative 107; Mismatches 188; Indels 45; Gaps 14;

Qy 8 PLKNHITPADIRALSKQLOQLADEVRGILTHTVTSISGHHFAAGLCTVELTVALHYFN 67
 Db 76 PLADTINVPFHKNLSVVELKQLSDELRSDFVIFHVSRTGHLSSLGVELTVALHYFN 135
 Qy 68 TPVDOLVWDVGHQAVPHKILITGRKERMPITRTIGVSAFPARDESEYDAFGVGHSSSTIS 127
 Db 136 APQDKIIMDVGHQSYFHKILITGRKRMPTMOTNLSGFTKRGSEHDCFGTGHSSSTIS 195
 Qy 128 AALGMAIASQLRSGEDKKMAVITIGDSITGMAVEAMNAGDVANANLVTILNDN-DMSI- 185
 Db 196 AGLGMAVAGDLKGNKNVAVITIGDAMTAGQAVEAMNAGVLDSDMVIILNDNKQVSLPT 255
 Qy 186 -----SPVGMANNVITKLSKFFYSVREESKXALAKMPS-VNELARKTEBHVKGMI- 236
 Db 256 ANLDGPPIPVGALSSALSKQSNRPLRREIRVAKGVTKKGGPHEQLAAKVDETARKMTIS 315
 Qy 237 -VPGTLFEELGFNYFGFIDGHEMLVSTLENKD-LTGPVFLAVVTKKKGAYAPAEKD 293
 Db 316 GSGSLFEELGLVYIGVDSHNIDLVSLINEVKSRTTGPVLHVTEKRGVPAVERA 375
 Qy 294 PLAYHGVPADEPT--KDFLPKAPSPHPTTYEVFGKRLCMAADBERLGLITPMRBSGS 351
 Db 376 ADKHGVKVPDPTGRQF--KTTNKTSYTYFEBALVABAEVDKDVALLHAAMGGSTG 432
 Qy 352 LVEFSQKFPNRYPDVAIAEQHNAVTLAAGQACQAGAPVVAISTFLQSGYDOLIHDAVON 411
 Db 433 LNLFPGRFPTRCFDVGIAEQHNAVTFPAAGLAGEGLPFCALISSFMGRAYDQVVDVLDQK 492
 Qy 412 LKMLPALDRAGLVGDPGPTHAGAFDYSYMRCLPNMLIMAPADENECROMLTTGFO-HHGP 470

```

      493 LPVRAMRAGLGVADGPTHCAGFDVTNACLPNNVYVMA-PSEDAELFHVATMAAIDRRP 552
      471 ASVRYRGRGPGRAID-----PTLALIEIGKAVVHAGSRILAILAMGSMTPAVEAGK- 523
      553 SCFRYPGRNGIGVELLGNKKGIP---LEIGKGRILIEBERVALLGYGAVONCLAAASL 608
      524 ----OLGATVVMRFEVFPDQALVLELARTHDVFTVBERNVLAGAGSALNTELOAKVL 579
      609 LHHGRATVADARFCXKEDRLIRSLASHVLTVERGSI--GGFGSHVQVFMALDGLL 667
      580 -----MPVCNIGLPDRFEVQGSREELSLVGLDSKGLIATI 615
      668 DKLKKRPV---LPDCYIDHSGSPVDQLSAGLTPSHIATV 706
  
```

RESULT 12

```

US-09-857-556A-33
Sequence 33, Application US/09857556A
Patent No. 6558915
GENERAL INFORMATION:
APPLICANT: Rebecca B. Cahoon
APPLICANT: Sean J. Coughlan
APPLICANT: Yong Tao
APPLICANT: Zude Weng
APPLICANT: Mark E. Williams
TITLE OF INVENTION: Plant 1-Deoxy-Xyloiose 5-Phosphate Synthase
FILE REFERENCE: B01290
CURRENT APPLICATION NUMBER: US/09/857,556A
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 60/110,779
PRIOR FILING DATE: 1998-12-03
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Microsoft Office 97
SEQ ID NO 33
LENGTH: 719
TYPE: prt
ORGANISM: Capsicum annuum
US-09-857-556A-33
  
```

```

Query Match      43.5%; Score 1397.5; DB 4; Length 719;
Best Local Similarity 46.1%; Pred. No. 2e-131;
Matches 294; Conservative 114; Mismatches 193; Indels 37; Gaps 13;
  
```

```

      8 PLEKNIHTPADIRALSKDQLQDLADVEVGYLTHTVSISGCHFAAGLGTVELTVALHYFN 67
      74 PIVDTINYPIMHKRNLSTKELKQDLADLRSDITFNVSKTGHLGSSLGVBELTVALHYFN 133
      68 TPVDQLVMDVGHQAVPHKILTGKREMPITRTLGVSAPFPADESEYDAFGVGHSSSTIS 127
      134 APQDITLMDVGHQAVPHKILTGKREMPITRTLGVSAPFPADESEYDAFGVGHSSSTIS 193
      128 AALGMAIASOLRGEDKKVVAIIGDSITGMAVEAMNAGDVANLILVINDN--DMSIS- 185
      194 AGLGMAVVRDLKGRNNNYIAVIGDGMATAGQAYEAMNAGYLDMSMITYLINDNQVSLPT 253
      186 -----PPVGMANNVLTXYLSSKFTYSSVRESKALAMPS--VWELARKTEEHVKMI- 236
      254 ATLDPGPPVPGVALSLSRLOSNNRLEIRLREYAKVTKQIGGPMHELAAKDEYARQWIS 313
      237 -VPGTLFPELGNVYGPIDGDMVEMLVSTLEMLD--LTGVPFLHVYTKKGGYAPAKED 293
      314 GSGSTLFEELGTYIGPVDGNIDILISLKEVSKTKTTSVULHVTEKRGYPIYAKERA 373
      294 PLAYHGVADPPT--KDFLPKAAAPSPHPTYTEVFGRMLCMAAODERLLGITPARESGG 351
      374 ADKHGVAKFDPATGKQFKSGAKTQSYTTY--FAEALIAEABADKDIVAIIHAAWGGGTG 430
      352 LVESQKPNRFPVDALAEQAVTLAAGQACGAPVVAISTYLGQGYDOLLIDVALLON 411
      431 MWLFLRRPTTCFVDGIAEQAHTVFAAGLACBGLPFCALYSSFPQRAYDDVADVLOK 490
      412 LMDLFDLRAGLVGPDGPTHAGAFDYSTYRCIPNNLIMAPADENECRQML--TTGQGHGPA 470
  
```

```

      491 LPVRAMRAGLGVADGPTHCAGFDVTNACLPNNVYVMA-PSEDAELFHVATMAAIDRRP 550
      471 ASVRYRGRGPGRAIDPTLTA--LEIGRAVEHGRSRIAILAMGSMTPAVEA-----GK 523
      551 SCFRYPGRNGIGVELLGNKKGIPLEVGKGRILIEBERVALLGYGAVONCLAAASVLSR 610
      524 OLGATVVMRFEVFPDQALVLELARTHDVFTVBERNVLAGAGSALNTELOAKVL----- 579
      611 GLQYTVADARFCXKEDRLIRSLASHVLTVERGSI--GGFGSHVQVFMALDGLDGL 669
      580 --MPVCNIGLPDRFEVQGSREELSLVGLDSKGLIATI 615
      670 KMRPVI---LPDRYIDHSGSPVDQLSAGLTPSHIATV 704
  
```

RESULT 13

```

US-09-146-221-4
Sequence 4, Application US/09146221
Patent No. 6190895
GENERAL INFORMATION:
APPLICANT: Croceau, Rodney B
APPLICANT: Lange, Bernd M
APPLICANT: Wildung, Mark R
APPLICANT: McCaskill, David G
TITLE OF INVENTION: Nucleic and Amino Acid Sequences Relating to a No. 6190895el
FILE REFERENCE: No. 6190895el transketolase and Methods for the Expression thereof
CURRENT APPLICATION NUMBER: US/09/146,221
CURRENT FILING DATE: 1998-09-01
EARLIER APPLICATION NUMBER: 60/056,033
EARLIER FILING DATE: 1997-09-02
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 721
TYPE: prt
ORGANISM: Mentha piperita
US-09-146-221-4
  
```

```

Query Match      43.3%; Score 1392; DB 3; Length 721;
Best Local Similarity 46.5%; Pred. No. 7.2e-131;
Matches 296; Conservative 115; Mismatches 189; Indels 36; Gaps 14;
  
```

```

      8 PLEKNIHTPADIRALSKDQLQDLADVEVGYLTHTVSISGCHFAAGLGTVELTVALHYFN 67
      77 PIVDTINYPIMHKRNLSTKELKQDLADLRSDITFNVSKTGHLGSSLGVBELTVALHYFN 136
      68 TPVDQLVMDVGHQAVPHKILTGKREMPITRTLGVSAPFPADESEYDAFGVGHSSSTIS 127
      137 TPDDKITMDVGHQAVPHKILTGKREMPITRTLGVSAPFPADESEYDAFGVGHSSSTIS 196
      128 AALGMAIASOLRGEDKKVVAIIGDSITGMAVEAMNAGDVANLILVINDN--DMSI- 184
      197 AGLGMAVVRDLKGRNNNYIAVIGDGMATAGQAYEAMNAGYLDMSMITYLINDNQVSLPT 256
      185 -----SPVGMANNVLTXYLSSKFTYSSVRESKALAMPS--VWELARKTEEHVKMI- 236
      257 ATVDGPAPVPGVALSKLTTLQASRKFQRLREBAKSMTKGAGAHBAHSLTLTVKGMWG 316
      237 VPG--TLPEHGFVYFGIDGDMVEMLVSTLEMLD--TPVPFLHVYTKKGGYAPAKED 293
      317 KPGASLFEELGTYIGV--DVEDLVYIFKVKEMPPAPVLIHITTEKRGYPIYAKERA 373
      294 PLAYHGVADPPTKDFLPKAAAPSPHPTYTEVFGRMLCMAAODERLLGITPARESGGLV 351
      374 ADKHGVAKFDPATGKQFKSGAKTQSYTTY--FAEALIAEABADKDIVAIIHAAWGGGTG 432
      354 EESQKPNRFPVDALAEQAVTLAAGQACGAPVVAISTYLGQGYDOLLIDVALLON 413
      433 IFQKQFPDRCTFDGIAEQAHTVFAAGLACBGLPFCALYSSFLQRYGQYDVADVLOK 492
      414 MLFDLRAGLVGPDGPTHAGAFDYSTYRCIPNNLIMAPADENECRQML--TTGQGHGPA 472
  
```



```

Db      493 VAPMDRAGVADGTHGADPTTMACLPNNVWAPDEDELAMITAAIIDR89C 552
Qy      473 VAYPRKGGAAIDPTL--TALEIGKAEVREHSGRIAILAMGSMVTPAVEAGKQI----- 525
Db      553 VAYPRKGGAAIDPTL--TALEIGKAEVREHSGRIAILAMGSMVTPAVEAGKQI----- 525
Qy      526 GATVVMREVRKPPDQALVLELARTHDVFTVEENVYAGAGSAINTELQACVY----- 579
Db      613 SYTVADARCKPDLGDLIKKLVOEHEVLITVEBGI--GGPSAHISHFSLNGLDGNLKM 671
Qy      580 MBVCNIGLPDRFEVCGREELSLVGLDSKGIAT 615
Db      672 RMVW---LPDRYIDHGAQSDQIEBAGLSFKHAGTV 704

```

RESULT 14
US-09-198-452A-1135
Sequence 1135, Application US/09198452A
Patent No. 6559294

GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 1135
LENGTH: 644
TYPE: PR1
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-1135

Query Match 35.7%; Score 1148; DB 4; Length 644;
Best Local Similarity 40.2%; Pred. No. 2.1e-106;
Matches 249; Conservative 115; Mismatches 236; Indels 20; Gaps 8;

```

Qy      8 PLKKNHTPADIRALSKDQLOGLADEVRGYLHTHTYISGGHFAAGLGYELTVLAHYVNT 67
Db      7 PLIDILSRADLKLSISQPLGAIEIRIRIISVLSQTHGLSSNLGIVELTIALHYVNS 66
Qy      68 TPVDQLVMDVGHQAYPHKILTGR--KERMTITLGGVSAFPARDESEYDAFGVSHSTSI 126
Db      67 SPDKXIFIVGHQTYPHKILTGRNNGFPHINDNGLSGFTNPTSDHDLFPSSAGATL 126
Qy      127 SAAIGMAISQURGEBKQVAILIGDSITGGMAYEMAKGVNANLVIYLANDNMSISP 186
Db      127 SAAIGMAISQURGEBKQVAILIGDSITGGMAYEMAKGVNANLVIYLANDNMSISP 186
Qy      187 PVGANNNTLTKVLSKFVSSVREESKKALAKP---SWMELARKTEBHVKGMTVGLTF 242
Db      186 NYGAMSRISRWLHNTKLTQVYKEMAKLIPRYDSLAKSRSLQCVKTLFCTPLTF 245
Qy      243 EELGENYFGPIDGHVEMLVSTLENKDLVGPVFLHVTYKKGATAPAEKDPVLAHYGVA 302
Db      246 EFGGLAYVGPIDGHVEMLVSTLENKDLVGPVFLHVTYKKGATAPAEKDPVLAHYGVA 305
Qy      303 -FPPTDPLPKAPSPHPTYTEVGRKMLDQMAAQDBRLGIPPARBGGGLVFPBQKPPN 361
Db      306 NPKKESAGHLPAIKKPSFPDIFGQTLCELGVSRLHVVTPASISRLGCFQKKEPE 365
Qy      362 RYFDVLAIEQAVTLAAGACOGAKFVAIYSTFLQGYDOLIHVALONLDMELFALDBA 421
Db      366 RYFDVLAIEQAVTLAAGACOGAKFVAIYSTFLQGYDOLIHVALONLDMELFALDBA 425
Qy      422 GLVGPQCPHAGKFDYSVRCIPKMLIMAPADENECQMLTGFQHNHGAASRYRKGKP 481
Db      426 GLVGPQCPHAGKFDYSVRCIPKMLIMAPADENECQMLTGFQHNHGAASRYRKGKP 485
Qy      482 GAALIDPLTA-----LEIGKAEVREHSGRIAILAMGSMVTPAVEAGKQI-----GATVVM 531

```

```

Db      486 HG--DP-LTGDPNFLRSGNAETLSQGBDYIILALGTLCTALSIKHQOLAYGISATVWD 542
Qy      532 MRVYPPDQALVLELARTHDVFTVEENVYAGAGSAINTELQACVY----- 591
Db      543 PIFKPPDNDLFFSLILMSHSHKVIITIBHSIRGHSSEFNVAATFNKVDILNPAIDTF 602
Qy      592 VEOGSEELSLVGLDSKI 611
Db      603 LSHGSKALTSIGLDESSK 622

```

RESULT 15
US-08-311-731A-56
Sequence 56, Application US/08311731A
Patent No. 6583266

GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESS: WOLF, GREENFIELD & SACKS, P. C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 736 amino acids
TYPES: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIUM LAPRAE
US-08-311-731A-56

Query Match 33.8%; Score 1086.5; DB 4; Length 736;
Best Local Similarity 38.1%; Pred. No. 4.1e-100;
Matches 237; Conservative 119; Mismatches 245; Indels 21; Gaps 9;

```

Qy      9 LKKNHTPADIRALSKDQLOGLADEVRGYLHTHTYISGGHFAAGLGYELTVLAHYVNT 68
Db      94 MLBOIRRRADLQHSQOGLRDLALIEBLVHKVAAAGGHLGPNVGLVETLTLARVDS 153
Qy      69 PVDDLVMDVGHQAYPHKILTGRKERMTITLGGVSAFPARDESEYDAFGVSHSTSI 128
Db      154 PHDPIPTGTHQAYVHKILTGRQDPDLRKKAGLGGVPSRAESHDVWESSHSTALSY 213
Qy      129 ALGMAISQURGEBKQVAILIGDSITGGMAYEMAKGVNANLVIYLANDNMSISP 187
Db      214 ADGLAKAFELAGNNRNVAVVAGDALTGCMCWALNNIATPRPVYIVANDNGRYAFT 273
Qy      188 VGANNNTLTKVLSKFVSSVREESKKALAKP---SWMELARKTEBHVKGMTVGLTFE 243

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 18.2424 Seconds
(without alignments) 3268.453 Million cell updates/sec

Title: US-09-941-947a-6

Perfect score: 3216

Sequence: 1 MKLTIDYPLAKNIHHPADIR.....LSVGLSDSKGLIATIEQFCA 620

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2097.5	65.2	626	2 H82266	1-deoxyxyulose-5-
2	2069.5	64.4	627	2 G83139	1-deoxyxyulose-5-
3	2015	62.7	619	2 A10385	1-deoxy-D-xyulose
4	2014.5	62.6	620	2 AF0554	1-deoxyxyulose-5-
5	2000.5	62.2	620	2 D64771	dxs protein - Esch
6	1995.5	62.0	620	2 B90688	1-deoxy-D-xyulose
7	1995.5	62.0	620	2 F85388	1-deoxy-D-xyulose
8	1951	60.7	620	2 A82582	deoxyxyulose-5-ph
9	1937.5	60.2	625	2 B64172	dxs protein - Haem
10	1776	55.2	637	2 B81978	probable 1-deoxyxy
11	1775	55.2	637	2 D81034	1-deoxyxyulose-5-
12	1667	51.8	608	2 A84984	dxs protein (impor
13	1620.5	49.5	628	2 A70376	conserved hypothet
14	1591.5	49.5	643	2 AD3376	1-deoxyxyulose-5-
15	1523.5	47.4	640	2 C87505	1-deoxyxyulose-5-
16	1499.5	46.6	639	2 A97450	1-deoxy-D-xyulose
17	1499.5	46.6	639	2 AC2668	1-deoxy-D-xyulose
18	1459	45.4	629	2 C83997	1-deoxyxyulose-5-
19	1452	44.8	629	2 S75175	hypothetical prote
20	1434	44.6	635	2 AF1881	1-deoxy-xyulose 5
21	1428	44.4	594	2 T02208	transketolase-like
22	1411.5	43.9	717	2 H85171	DBF (CIA1) protei
23	1410.5	43.9	641	2 G28771	hypothetical prote
24	1402.5	43.6	739	2 T52289	probable transke
25	1402	43.6	739	2 T08140	1-deoxy-D-xyulose
26	1400	43.5	632	2 A11607	D-1-deoxyxyulose
27	1397.5	43.5	719	2 T09543	deoxyxyulose synt
28	1394.5	43.4	633	2 B69961	conserved hypothet
29	1366.5	42.5	619	2 A97156	deoxyxyulose-5-ph

30	1351	42.0	609	2 AE1245	D-1-deoxyxyulose
31	1326	41.2	629	2 G75390	1-deoxy-D-xyulose
32	1274.5	39.6	608	2 A72213	1-deoxyxyulose-5-
33	1259.5	39.2	703	2 D71420	hypothetical prote
34	1227.5	38.2	630	2 B71276	probable transke
35	1227	38.2	615	2 E81451	1-deoxyxyulose-5-
36	1148	35.7	644	2 A86623	transketolase (imp
37	1148	35.7	644	2 A72002	1-deoxyxyulose-5-
38	1119.5	34.8	656	2 T35408	probable transke
39	1106.5	34.4	618	2 H71946	1-deoxyxyulose-5-
40	1099.5	34.2	618	2 B64564	transketolase B -
41	1087.5	33.8	638	2 E70528	probable dxs prote
42	1086.5	33.8	643	2 H87038	hypothetical prote
43	1072.5	33.3	632	2 F81684	1-deoxyxyulose-5-
44	1044.5	32.5	640	2 F71527	probable transke
45	886	27.5	580	2 A68354	1-deoxyxyulose-5-

ALIGNMENTS

RESULT 1

H82266 1-deoxyxyulose-5-phosphate synthase VC0889 [imported] - Vibrio cholerae (strain N16961)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: H82266

R:Heidelberg, J.F.; Bisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.B.; Dodson, R.J.;

cardson, D.; Ermolaeva, M.D.; Vamathevan, U.; Bass, S.; Qin, H.; Dragoti, I.; Sellers, P.

1. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.W.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: H82266

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-626 <H81>

A:Cross-references: GB:AE004173; GB:AE003852; NID:59655341; PID:AAF94051.1; GSPDB:GN001;

A:Experimental source: serogroup O1, strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0889

A:Map position: 1

C:Superfamily: hypothetical protein C2814

Query Match	Score	Length	DB 2;	Length	DB 2;
Best Local Similarity	62.5%	Pred. No. 2.8e-139;			
Matches	391;	Conservative 109;	Mismatches 117;	Indels	9;
				Gaps	3;
QY	3	LTLD---YPLAKNIHTPADIRALSKDQLQCLADVRGVLTTTSGGHPAAGTVELT	59		
DB	1	MTLDISKYPTLALANTPDELRLPEVLPKLCDELRITLINSVSQSSCHLASGLTVELT	60		
QY	60	VALATVFTPTVDQVWDVGHQAYPHKILTGKREMERITRTIGVSAPFAPDESBYDARGV	119		
DB	61	VALATVFTPTDLITVDYGHQAYPHKILTGKREMERITRTIGVSAPFAPDESBYDARGV	120		
QY	120	GHSSTISAAIGMAIASQLRBDKXKVAIIDGSIIGGMAEYANNEADYVNNALLVIND	179		
DB	121	GHSSTISAAIGMAIASQLRBDKXKVAIIDGSIIGGMAEYANNEADYVNNALLVIND	180		
QY	180	NDMGSIPEVGMANNVLTIVLSKFFSVREESKALALAPGPEVIELARTSEIVKQIVPG	239		
DB	181	NEMGISSEVGMANNVLTIVLSKFFSVREESKALALAPGPEVIELARTSEIVKQIVPG	240		
QY	240	TLFEELGNYRQPIDGHDVEMVSTLENLKDVLTPGVFLHVVTKKGGYAPAEKDFLAHVG	299		
DB	241	TLFEELGNYRQPIDGHDVEMVSTLENLKDVLTPGVFLHVVTKKGGYAPAEKDFLAHVG	300		
QY	300	VPAFPTDQFLPKAASPHPTYTEVFGRLCDMAAODERLGITFPAEAGSGLVESQKF	359		
DB	301	VPAFPTDQFLPKAASPHPTYTEVFGRLCDMAAODERLGITFPAEAGSGLVESQKF	359		
QY	360	PNRYEDVAIAEQAHTLAAGAACQAKPVAIYSTFELRGVDLIHDVALQNLDMFLALD	419		

```

Db      360 PQQYEVVAIAEQAHTVLAATGMAIAGHPVAIYSTFLOQGYDOLIHDAVIMPLPMAFID 419
Qy      420 RAGLVDPDGPZTHAGADYSYMRCTIPNMLIMAPADENECQMLTTGQHHGPASVAPYPAK 479
Db      420 RAGIVADQGTTHAGADYSYMRCTIPNMLIMAPADENECQMLTTGQHHGPASVAPYPAK 479
Qy      480 GPCALIDPTLFAIEIGKAEVRH-----HGSRJATIAMGSMWTPAYEAGKQATVNNRF 534
Db      480 GMEVLESSFTALIEIGKRLMERSTACEGEKVAIISFTLLPNAIQAKMAATVADRKF 539
Qy      535 VKPEPDALVLELARTDVFVTEENVIAGGASALNTFQAOKYMLPVCNIGLPRPFRFQ 594
Db      540 VKPELDELKQIAQHTDVTLEENVIAGGAGVIEFLMERKQIKPVNLGLPDPQFVQ 599
Qy      595 GSRBELLSLVGLDSKGIATITBQFCA 620
Db      600 GTOEWHALGLDGAIGIERAIDVLA 625

```

RESULT 2

1-deoxy-5-phosphate synthase PA4044 [imported] - *Pseudomonas aeruginosa* (strain G83139)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: G83139
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; Lim, J.; Lory, S.; Olson, M.V.
 Mature 406, 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
 A:Reference number: AB2950; MUID:2043737; PMID:1096043
 A:Accession: G83139
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-627 <STO>
 A:Cross-References: GB:AE004821; GB:AE004091; NID:99950236; PIDN:AA07431.1; GSPDB:GN001
 A:Experimental source: strain PA01
 A:Genetics:
 A:Gene: dks; PA4044
 C:Superfamily: hypothetical protein C2814

Query Match 64.4%; Score 2069.5; DB 2; Length 627;
 Best Local Similarity 63.4%; Pred. No. 2.6e-137;
 Matches 388; Conservative 100; Mismatches 117; Indels 7; Gaps 2;

```

Qy      8 PILLKNIHPADIRALSKQLOQLADEVRGVLHTVTSISGCHFAAGLGVETLVLAHYVN 67
Db      16 PLMDASSFAELRRLEADLETLADELRQYLLVIGQTHGFGAGLGVETLVLAHYVD 75
Qy      68 TEVDQLVMDVGHQAVPHKILTGKERMPTIRTLGVSAFPADESEYDAFGVGHSTSI 127
Db      76 TEDDRLVMDVGHQAVPHKILTERRELMTLRQKGLAEPRADESEYDTFGVGHSTSI 135
Qy      128 PALLGAIASQLGDEKQKVAIIGDGSITGMAIYEMNNAAGVNNALVTLINDNMSISPP 187
Db      136 PALLGAIASQLGDEKQKVAIIGDGSITGMAIYEMNNAAGVNNALVTLINDNMSISPP 195
Qy      188 VGMANNVTLKVLSSKFSYSSVREESKKALAKMPSVMEIARKTEEHVKGMIIVPTLFEELG 247
Db      196 VGLLSNVLAKILSSRTYSRMSRSGKKVLSRLPGAMEIARRTEHYAKMGLVPTLFEELG 255
Qy      248 NYFGPIDGHVEMVSTLENLKDLPVPLHVTYTKGKGYAPAEKDPPLAHGVPAEDPTX 307
Db      256 NYFGPIDGHVEMVSTLENLKDLPVPLHVTYTKGKGYAPAEKDPPLAHGVPAEDPTX 310
Qy      308 DFLPAPAP--SPHPTTVEVFGRLCDMAAODERLLGITPPARESGGLVEFSQKPNRYFD 365
Db      311 LEAPSSAPKRTGSGPYSSVFGQMLCDMAAODARLLGITPPARESGGLVEFSQKPNRYFD 370
Qy      366 VAIAGQHAVTTLAAGQACGAGPVVAIYSTFLOQGYDOLIHDAVIMPLPMAFID 425
Db      371 VAIAGQHAVTTLAAGQACGAGPVVAIYSTFLOQGYDOLIHDAVIMPLPMAFID 430

```

```

Qy      426 PDGTHAGADYSYMRCTIPNMLIMAPADENECQMLTTGQHHGPASVAPYPAK 485
Db      431 EDGTHAGADYSYMRCTIPNMLIMAPADENECQMLTTGQHHGPASVAPYPAK 490
Qy      486 DPTLTAIEIGKAEVRH-----HGSRJATIAMGSMWTPAYEAGKQATVNNRF 545
Db      491 DPTLTAIEIGKAEVRH-----HGSRJATIAMGSMWTPAYEAGKQATVNNRF 550
Qy      546 LARTDVFVTEENVIAGGASALNTFQAOKYMLPVCNIGLPRPFRFQ 605
Db      551 LAGSHELLVTEENVIAGGASALNTFQAOKYMLPVCNIGLPRPFRFQ 610
Qy      606 LDSKGIATITBQFCA 617
Db      611 LDMAGIERAVRQ 622

```

RESULT 3

1-deoxy-D-xylulose 5-phosphate synthase (EC 2.2.-.-) [imported] - *Yersinia pestis* (strain A10385)
 C:Species: *Yersinia pestis*
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
 C:Accession: A10385
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; H., M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, N.
 Mature 413, 523-527, 2001
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: A10385
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-619 <KIR>
 A:Cross-References: GB:AF590842; PIDN:CA92412.1; PID:91598115; GSPDB:GN00175
 A:Genetics:
 A:Gene: dks
 C:Superfamily: hypothetical protein C2814
 C:Keywords: transferase

Query Match 62.7%; Score 2015; DB 2; Length 619;
 Best Local Similarity 61.6%; Pred. No. 1.7e-133;
 Matches 378; Conservative 95; Mismatches 139; Indels 2; Gaps 2;

```

Qy      7 YPLKNIHPADIRALSKQLOQLADEVRGVLHTVTSISGCHFAAGLGVETLVLAHYVN 66
Db      8 YPLALANFBEELRMLPKESLPKLCBRLQYLLVIGQTHGFGAGLGVETLVLAHYVD 67
Qy      67 NTPVDQLVMDVGHQAVPHKILTGKERMPTIRTLGVSAFPADESEYDAFGVGHSTSI 126
Db      68 NTPVDQLVMDVGHQAVPHKILTGKERMPTIRTLGVSAFPADESEYDAFGVGHSTSI 127
Qy      127 SAALGMAIASQLGDEKQKVAIIGDGSITGMAIYEMNNAAGVNNALVTLINDNMSISPP 186
Db      128 SAALGMAIASQLGDEKQKVAIIGDGSITGMAIYEMNNAAGVNNALVTLINDNMSISPP 187
Qy      187 PVGMANNVTLKVLSSKFSYSSVREESKKALAKMPSVMEIARKTEEHVKGMIIVPTLFEELG 246
Db      188 PVGMANNVTLKVLSSKFSYSSVREESKKALAKMPSVMEIARKTEEHVKGMIIVPTLFEELG 247
Qy      247 FNYFGPIDGHVEMVSTLENLKDLPVPLHVTYTKGKGYAPAEKDPPLAHGVPAEDPTX 306
Db      248 FNYFGPIDGHVEMVSTLENLKDLPVPLHVTYTKGKGYAPAEKDPPLAHGVPAEDPTX 307
Qy      307 KDPLPKAPSPHPTTVEVFGRLCDMAAODERLLGITPPARESGGLVEFSQKPNRYFD 366
Db      308 SGLTLPK--SQSLPTLYSKIFGEMLCETPAKDSKIMAVTPARESGGLVEFSQKPNRYFD 366
Qy      367 AIAEQHAVTTLAAGQACGAGPVVAIYSTFLOQGYDOLIHDAVIMPLPMAFID 426
Db      367 AIAEQHAVTTLAAGQACGAGPVVAIYSTFLOQGYDOLIHDAVIMPLPMAFID 426
Qy      427 DGTHAGADYSYMRCTIPNMLIMAPADENECQMLTTGQHHGPASVAPYPAK 486

```

Db 427 DQTHQAFDLSIFKCIPIIMVIMAPSDENECRQMLFTGHNHNGPAAYVPRNGKSAVLE 486
 487 PTLTALEIGKAEVRHHSRIALAMGSMTVPAYEAGKQAGATVVMNRFPKPDQALVLE 546
 487 P-LEMLPIGKGLVRBEGKALICTCTLLAQKQALAEENLNTLVNRRVKKPLDESLVLEM 545
 547 ARTHDVFVTEENVNLAGGASAINTEFLQAKVLMFVNCIGLPDRFEQSRRELLSLVGL 606
 546 AAKHQLVLTVEENALIMGAGSGVNEELMAKRKRVPLVNLGLPDLFVPGQDEKMSSELGL 605
 Qy 607 DSKGILATIEQFCA 620
 Db 606 DAAGIQRIETAMLA 619

RESULT 4

AF0554
 1-deoxyxylulose-5-phosphate synthase [imported] - Salmonella enterica subsp. enterica sero-
 C:Species: Salmonella enterica subsp. enterica serovar Typhimurium
 A:Note: This species has also been called Salmonella typhimurium
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AF0554
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A:Author: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
 A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AF0554
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-620 <PAR>
 A:Cross-references: GB:AU513382; PIDN:CAD08878.1; PID:g16501690; GSPDB:GN00176
 C:Genetics: dks
 A:Gene: dks
 C:Superfamily: hypothetical protein C2814

Query Match 62.6%; Score 2014.5; DB 2; Length 620;
 Best Local Similarity 61.0%; Pred. No. 1.8e-133;
 Matches 375; Conservative 106; Mismatches 131; Indels 3; Gaps 3;

Qy 7 YPLAKNHTPADIRALSKDQLOQLADEVRGYLTHTVSISGSHFAGLGTVELTVLAHYV 66
 8 YPTLALVSTQETRLIPKESLPKCDLRLRYLDSVSSSGHFPASGLGTVELTVLAHYV 67
 Db 8 YPTLALVSTQETRLIPKESLPKCDLRLRYLDSVSSSGHFPASGLGTVELTVLAHYV 67
 Qy 67 NTPVDQLVMDVGHQAYPHKILTGKREMPITRTIGGSAFPAARDESEYDAFVGHSSSTSI 126
 68 NTPVDQLVMDVGHQAYPHKILTGKREMPITRTIGGSAFPAARDESEYDAFVGHSSSTSI 127
 Db 68 NTPVDQLVMDVGHQAYPHKILTGKREMPITRTIGGSAFPAARDESEYDAFVGHSSSTSI 127
 Qy 127 SAALGMAIASQLRGEDKKMAVITGDSITGMAVEANMHAGDVNANLTVILNDMSISP 186
 128 SAGIGIAVAAREGKORRTVCVIGDGLITAGMAFEANMHAGDIRPDMVLINDMMSISE 187
 Db 128 SAGIGIAVAAREGKORRTVCVIGDGLITAGMAFEANMHAGDIRPDMVLINDMMSISE 187
 Qy 187 PVGAMNNYLTLYLSSKREYSSVRESKKAALAKMPSVWELARKTEHYKGMIVPGTLFSELG 246
 188 NVGALNNHLLARLSSGKLYSSLRBGGKVFSGVPPIKELKRTBEHKKGMVPGTLFSELG 247
 Db 188 NVGALNNHLLARLSSGKLYSSLRBGGKVFSGVPPIKELKRTBEHKKGMVPGTLFSELG 247
 Qy 247 FNYPGPIGDHVDVMTLSTLKNRDLKGPFLHMTKKRGYEPKADPITTHAVPKFDPDS 307
 248 FNYPGPIGDHVDVMTLSTLKNRDLKGPFLHMTKKRGYEPKADPITTHAVPKFDPDS 307
 Db 248 FNYPGPIGDHVDVMTLSTLKNRDLKGPFLHMTKKRGYEPKADPITTHAVPKFDPDS 307
 Qy 307 KDFLPKAPSPHPTVTEFGRWLCDMAODERLIGITPARBEGSGLVESQKFPNRYEDV 366
 308 SGCLPKSS-GGLPGYSKIFGDLCTEANDSKLMAITPARBEGSGLVESQKFPNRYEDV 366
 Db 308 SGCLPKSS-GGLPGYSKIFGDLCTEANDSKLMAITPARBEGSGLVESQKFPNRYEDV 366
 Qy 367 AIAEGHATVTLAAGACCGAKPVVALYSTFQRGYQOLIHVALQNLDMTLPALDRAGLVGP 426
 367 AIAEGHATVTLAAGACCGAKPVVALYSTFQRGYQOLIHVALQNLDMTLPALDRAGLVGP 426
 Db 367 AIAEGHATVTLAAGACCGAKPVVALYSTFQRGYQOLIHVALQNLDMTLPALDRAGLVGP 426
 Qy 427 DQTHQAFDLSIFKCIPIIMVIMAPSDENECRQMLFTGHNHNGPAAYVPRNGKSAVLE 486
 427 DQTHQAFDLSIFKCIPIIMVIMAPSDENECRQMLFTGHNHNGPAAYVPRNGKSAVLE 486

Db 427 DQTHQAFDLSIFKCIPIIMVIMAPSDENECRQMLFTGHNHNGPAAYVPRNGKSAVLE 486
 487 PTLTALEIGKAEVRHHSRIALAMGSMTVPAYEAGKQAGATVVMNRFPKPDQALVLE 545
 487 TP-LEMLPIGKGLVRBEGKALICTCTLLAQKQALAEENLNTLVNRRVKKPLDESLVLEM 545
 546 ARTHDVFVTEENVNLAGGASAINTEFLQAKVLMFVNCIGLPDRFEQSRRELLSLVGL 605
 546 MAQHDLVLTVEENALIMGAGSGVNEELMAKRKRVPLVNLGLPDLFVPGQDEKMSSELGL 605
 Qy 606 DSKGILATIEQFCA 620
 Db 606 DAAGIQRIETAMLA 619

RESULT 5

D64771
 dks protein - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: D64771
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col, A.; Rose, D.U.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: D64771
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-620 <BALT>
 A:Cross-references: GB:AB000148; GB:U00096; NID:g1786614; PIDN:AACT3523.1; PID:g1786622;
 A:Experimental source: strain K-12, substrain M61655
 C:Genetics: dks
 A:Gene: dks
 C:Superfamily: hypothetical protein C2814

Query Match 62.2%; Score 2000.5; DB 2; Length 620;
 Best Local Similarity 60.2%; Pred. No. 1.8e-132;
 Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

Qy 7 YPLAKNHTPADIRALSKDQLOQLADEVRGYLTHTVSISGSHFAGLGTVELTVLAHYV 66
 8 YPTLALVSTQETRLIPKESLPKCDLRLRYLDSVSSSGHFPASGLGTVELTVLAHYV 67
 Db 8 YPTLALVSTQETRLIPKESLPKCDLRLRYLDSVSSSGHFPASGLGTVELTVLAHYV 67
 Qy 67 NTPVDQLVMDVGHQAYPHKILTGKREMPITRTIGGSAFPAARDESEYDAFVGHSSSTSI 126
 68 NTPVDQLVMDVGHQAYPHKILTGKREMPITRTIGGSAFPAARDESEYDAFVGHSSSTSI 127
 Db 68 NTPVDQLVMDVGHQAYPHKILTGKREMPITRTIGGSAFPAARDESEYDAFVGHSSSTSI 127
 Qy 127 SAALGMAIASQLRGEDKKMAVITGDSITGMAVEANMHAGDVNANLTVILNDMSISP 186
 128 SAGIGIAVAAREGKORRTVCVIGDGLITAGMAFEANMHAGDIRPDMVLINDMMSISE 187
 Db 128 SAGIGIAVAAREGKORRTVCVIGDGLITAGMAFEANMHAGDIRPDMVLINDMMSISE 187
 Qy 187 PVGAMNNYLTLYLSSKREYSSVRESKKAALAKMPSVWELARKTEHYKGMIVPGTLFSELG 246
 188 NVGALNNHLLARLSSGKLYSSLRBGGKVFSGVPPIKELKRTBEHKKGMVPGTLFSELG 247
 Db 188 NVGALNNHLLARLSSGKLYSSLRBGGKVFSGVPPIKELKRTBEHKKGMVPGTLFSELG 247
 Qy 247 FNYPGPIGDHVDVMTLSTLKNRDLKGPFLHMTKKRGYEPKADPITTHAVPKFDPDS 307
 248 FNYPGPIGDHVDVMTLSTLKNRDLKGPFLHMTKKRGYEPKADPITTHAVPKFDPDS 307
 Db 248 FNYPGPIGDHVDVMTLSTLKNRDLKGPFLHMTKKRGYEPKADPITTHAVPKFDPDS 307
 Qy 307 KDFLPKAPSPHPTVTEFGRWLCDMAODERLIGITPARBEGSGLVESQKFPNRYEDV 366
 308 SGCLPKSS-GGLPGYSKIFGDLCTEANDSKLMAITPARBEGSGLVESQKFPNRYEDV 366
 Db 308 SGCLPKSS-GGLPGYSKIFGDLCTEANDSKLMAITPARBEGSGLVESQKFPNRYEDV 366
 Qy 367 AIAEGHATVTLAAGACCGAKPVVALYSTFQRGYQOLIHVALQNLDMTLPALDRAGLVGP 426
 367 AIAEGHATVTLAAGACCGAKPVVALYSTFQRGYQOLIHVALQNLDMTLPALDRAGLVGP 426
 Db 367 AIAEGHATVTLAAGACCGAKPVVALYSTFQRGYQOLIHVALQNLDMTLPALDRAGLVGP 426
 Qy 427 DQTHQAFDLSIFKCIPIIMVIMAPSDENECRQMLFTGHNHNGPAAYVPRNGKSAVLE 486
 427 DQTHQAFDLSIFKCIPIIMVIMAPSDENECRQMLFTGHNHNGPAAYVPRNGKSAVLE 486
 Db 427 DQTHQAFDLSIFKCIPIIMVIMAPSDENECRQMLFTGHNHNGPAAYVPRNGKSAVLE 486
 Qy 486 DPTLTALEIGKAEVRHHSRIALAMGSMTVPAYEAGKQAGATVVMNRFPKPDQALVLE 545
 486 DPTLTALEIGKAEVRHHSRIALAMGSMTVPAYEAGKQAGATVVMNRFPKPDQALVLE 545

Db 487 TP-LKLPKIGKGVKRGKSKALINFGTILMPKAAKASLNTIVDMRVKRLDELALIE 545
 QY 546 LARHDVFTVEBNVNIAGAGSALNTFLQOKVLPVCNIGLPDRFVEQSRBELLSVG 605
 Db 546 MAASHALVTVEBNALMGAGSGVNEVLMAHRKRPVPLNIGLPDFIPQGTQEMRAELG 605
 QY 606 LDSKGLIATIEQCA 620
 Db 606 LDAAGMEAKIKAMLA 620

RESULT 6

B90688
 1-deoxy-D-xylose 5-phosphate synthase [imported] - Escherichia coli (strain O157:H7, s
 C.Species: Escherichia coli
 C.Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C.Accession: B90688
 R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
 DNA Res. 8, 11-22, 2001
 A.Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
 A.Reference number: A99629; UID:21156231; PMID:11258796
 A.Accession: B90688
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-620 <HAY>
 A.Cross-references: GB:BA00007; PIDN:BA03897.1; PID:G13359931; GSPDB:GN00154
 A.Experimental source: strain O157:H7, substrain RIMD 0509552
 C.Genetics:
 A.Gene: BCS0474
 C.Superfamily: hypothetical protein C2814

Query Match 62.0%; Score 1995.5; DB 2; Length 620;
 Best Local Similarity 59.8%; Pred. No. 4e-132;
 Matches 368; Conservative 115; Mismatches 129; Indels 3; Gaps 3;

QY 7 YPLKNIHPADIRALSKQLOQLADENVGYLITHTVISISGHPAAGLGTVELTVALHYF 66
 Db 8 YPLALVSTQELRLPKESLPKLCDELRLRYLDSVRSRSGHFASGLGTVELTVALHYV 67
 QY 67 NTPVDQVMDVGHQAVPHKILTGKERMPTITLIGVSAFPARDSESEYDAFGVGHSSISI 126
 Db 68 NTPDQLIMDVGHQAVPHKILTGKERMPTITLIGVSAFPARDSESEYDAFGVGHSSISI 127
 QY 127 SAALGMAIASQLRGEDEKKNVAIIIGDSITGMAVYEMNHAGDVANMLVITLNDMSISP 186
 Db 128 SAGIGIAVAAEKKGKRRVTCVIGDAITAGMAFEAMNHAGDIRPDMVLVNDNMSISE 187
 QY 187 PVGAMNNYITKVLSSKFYSVREESKKALAKMPSVVELARKTEEHVGMIVPGLFEELG 246
 Db 188 NGALNNHIAQLISGLYSILRSGKRVSGVPPIKELKRTEHHIKGVVPGTLFEELG 247
 QY 247 FNYFGPIIDGVEMLVSTLNTKDLTGVPFLHVVTKGKGVAAPAKDPLAHGVPAEDPT 306
 Db 248 FNYIGVDGHDVGLITLTKNMRDLKGPFLHMTKKGKRGVEPAEDPTITFAVVKFDP 307
 QY 307 KQFLPKAAPSHPPTTEVFGKMLCMAAODERLIGITPAMREGSGLVESQKFPARYPDV 366
 Db 308 SGCLPKSS-GGLPYSKIRFDMLCETAAKDKMLAITPAMREGSGVNEVSRKFPARYDV 366
 QY 367 AIAEOHVAITLAAGACGAKPVVAIYSTFLORGQDLIHVALONLMDLPAIDRAGLVGP 426
 Db 367 AIAEOHVAITPAAGALIGYKPIVAIYSTFLORAYDQVLIHVALIKLVLPAIDRAGLVGA 426
 QY 427 DGPTHAGADYSYKRCIPNMLIAPADENRCQMLTTGFOHN-GRASVYPRGKPGAAI 485
 Db 427 DQGTGAGADLSYKRCIPENVMITPSENCRCQMLTYGHVNDGSARVPRGKPGAVGEL 486
 QY 486 DPTTLALTEIGKAEVRHSGRIALAMGSNVTVAEAKGOLGATVNNRFPVFPDQALVLE 545
 Db 487 TP-LKLPKIGKGVKRGKSKALINFGTILMPKAAKASLNTIVDMRVKRLDELALIE 545

QY 546 LARHDVFTVEBNVNIAGAGSALNTFLQOKVLPVCNIGLPDRFVEQSRBELLSVG 605
 Db 546 MAASHALVTVEBNALMGAGSGVNEVLMAHRKRPVPLNIGLPDFIPQGTQEMRAELG 605
 QY 606 LDSKGLIATIEQCA 620
 Db 606 LDAAGMEAKIKAMLA 620

RESULT 7

F85338
 1-deoxy-D-xylose 5-phosphate synthase [imported] - Escherichia coli (strain O157:H7, s
 C.Species: Escherichia coli
 C.Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 27-Nov-2001
 C.Accession: F85338
 R.Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.O.; Davis, N.W.; Lim, A.; Dialante, E.; Potamousis, K.; Apodaca,
 Nature 409, 525-533, 2001
 A.Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A.Reference number: A85480; UID:21074935; PMID:11206551
 A.Accession: F85338
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-620 <STO>
 A.Cross-references: GB:AE005174; UID:G12513276; PIDN:AG54770.1; GSPDB:GN00145; UWGP:205
 A.Experimental source: strain O157:H7, substrain EDL933
 C.Genetics:
 A.Gene: dks
 C.Superfamily: hypothetical protein C2814

Query Match 62.0%; Score 1995.5; DB 2; Length 620;
 Best Local Similarity 59.8%; Pred. No. 4e-132;
 Matches 368; Conservative 115; Mismatches 129; Indels 3; Gaps 3;

QY 7 YPLKNIHPADIRALSKQLOQLADENVGYLITHTVISISGHPAAGLGTVELTVALHYF 66
 Db 8 YPLALVSTQELRLPKESLPKLCDELRLRYLDSVRSRSGHFASGLGTVELTVALHYV 67
 QY 67 NTPVDQVMDVGHQAVPHKILTGKERMPTITLIGVSAFPARDSESEYDAFGVGHSSISI 126
 Db 68 NTPDQLIMDVGHQAVPHKILTGKERMPTITLIGVSAFPARDSESEYDAFGVGHSSISI 127
 QY 127 SAALGMAIASQLRGEDEKKNVAIIIGDSITGMAVYEMNHAGDVANMLVITLNDMSISP 186
 Db 128 SAGIGIAVAAEKKGKRRVTCVIGDAITAGMAFEAMNHAGDIRPDMVLVNDNMSISE 187
 QY 187 PVGAMNNYITKVLSSKFYSVREESKKALAKMPSVVELARKTEEHVGMIVPGLFEELG 246
 Db 188 NGALNNHIAQLISGLYSILRSGKRVSGVPPIKELKRTEHHIKGVVPGTLFEELG 247
 QY 247 FNYFGPIIDGVEMLVSTLNTKDLTGVPFLHVVTKGKGVAAPAKDPLAHGVPAEDPT 306
 Db 248 FNYIGVDGHDVGLITLTKNMRDLKGPFLHMTKKGKRGVEPAEDPTITFAVVKFDP 307
 QY 307 KQFLPKAAPSHPPTTEVFGKMLCMAAODERLIGITPAMREGSGLVESQKFPARYPDV 366
 Db 308 SGCLPKSS-GGLPYSKIRFDMLCETAAKDKMLAITPAMREGSGVNEVSRKFPARYDV 366
 QY 367 AIAEOHVAITLAAGACGAKPVVAIYSTFLORGQDLIHVALONLMDLPAIDRAGLVGP 426
 Db 367 AIAEOHVAITPAAGALIGYKPIVAIYSTFLORAYDQVLIHVALIKLVLPAIDRAGLVGA 426
 QY 427 DGPTHAGADYSYKRCIPNMLIAPADENRCQMLTTGFOHN-GRASVYPRGKPGAAI 485
 Db 427 DQGTGAGADLSYKRCIPENVMITPSENCRCQMLTYGHVNDGSARVPRGKPGAVGEL 486
 QY 486 DPTTLALTEIGKAEVRHSGRIALAMGSNVTVAEAKGOLGATVNNRFPVFPDQALVLE 545
 Db 487 TP-LKLPKIGKGVKRGKSKALINFGTILMPKAAKASLNTIVDMRVKRLDELALIE 545
 QY 546 LARHDVFTVEBNVNIAGAGSALNTFLQOKVLPVCNIGLPDRFVEQSRBELLSVG 605
 Db 546 MAASHALVTVEBNALMGAGSGVNEVLMAHRKRPVPLNIGLPDFIPQGTQEMRAELG 605

QY 606 LDSKILATIEQFCA 620
 DB 606 LDAGGAKIKAMLA 620

RESULT 8

A25582
 deoxyxylulose-5-phosphate synthase XF2249 (imported) - Xylella fastidiosa (strain 9asc)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: A25582
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A25515; PMID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: A25582
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-670 <SIM>
 A:Cross-references: GB:AEO04037, GB:AEO03849, NID:g9107394, PIDN:AAE95048.1; GSPDB:GN001
 A:Experimental source: strain 9asc
 R:Simson, A.J.G.; Relinack, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrero, D.M.; Carreir, H
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.B.; Kurmae, B.E.; Laig
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.P.; Martino, C.L.; Marques, M.V.; Martins, B
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiter, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak
 M.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tsuchiko, M.R.; Vallada, H.; Van Sluys, W.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 A:Gene: XF2249
 C:Superfamily: hypothetical protein C2814

Query Match 60.7%; Score 1951; DB 2; Length 670;
 Best Local Similarity 59.3%; Pred. No. 5,9e-129;
 Matches 365; Conservative 98; Mismatches 147; Indels 6; Gaps 3;

4 TTDYPLKNIHTPADRALSKDLOOLADEVGYLTHTVTSISGHPAAGLGVETLVATLH 63
 DB 39 STCYPRLSIQIPEDLTQTESELAADVADLANLYLESVLSGHPAAGLGVETLVATLH 98
 QY 64 YFNTPTVDOLVMDVGHQAVPHKILTGKRRMPTITLGGVSAFPARDESEYDAFVGHSS 123
 DB 99 YVCTPPIIDLWMDVGHQAVPHKILTGKRRMPTITLGGVSAFPARDESEYDAFVGHSS 158
 QY 124 TGISALGKAIASOLGEBDKKVAIIIGDSITGGMAVEAMNHAGVNV--ANLTVLTIND 181
 DB 159 TGISALGKAIASOLGEBDKKVAIIIGDSITGGMAVEAMNHAGVNV--ANLTVLTIND 218
 QY 182 MSISPPVGMANNLYTKVLSKFEYSVRESKCALA--KAPSVWELAKTEETHVKMIVPG 239
 DB 219 MSISEAVGGLTLMGLATGSKALNMRGSKLIPGCKTNATATRLRREETHVKMIVPG 278
 QY 240 TLFEELGPRVPIGDHVENMLVSTLENLKDITGPVFLHVTTKKGYAPAEKDLAHHG 299
 DB 279 TLFEELGPRVPIGDHVENMLVSTLENLKDITGPVFLHVTTKKGYAPAEKDLAHHG 338
 QY 360 VPAPPTKQFLPKAAPSHPITVTFVGRWLCDMAAODERLLGTPAMREGSGLVFESQKF 359
 DB 339 VAPFPQKGLIYAGAKK--QYTVDFSEWLCMAAVEPRLTATPAMEGSGLVFESQKF 396
 QY 360 PRRYVDVAIAEGHATVLAAGACGAKPVVAIYSTFLQSGYQLHVDVALQNLDMFLD 419
 DB 397 PRRYVDVAIAEGHATVLAAGACGAKPVVAIYSTFLQSGYQLHVDVALQNLDMFLD 456

QY 420 PAILVGPDPGPTHAGAFDYISYKRCIPNMLTMAPADENECROKLTTPQHHGPASVYPRGK 479
 DB 457 RGVVGPDPGPTHAGAFDYISYKRCIPNMLTMAPADENECROKLTTPQHHGPASVYPRGK 516
 QY 480 GPGAIIDPTLALTEIGKREVENHHSRIATLAWGSMVTPAVYAGKQIAGTVNMVFVKPD 539
 DB 517 GPGVPSASLKVLPVGAQALHSGSTRILALGFCVAPAEQVGRRLTLVNMVFVKPD 576
 QY 540 QALVELARTHDVFTVYEVNVIAGAGSALNTPLQAKVLMVPCNIGLPDRFVGGSRRE 599
 DB 577 RTLLLEARTHEGFTVTEEDNVVAGAGSVALINAEGLVPIVHLGLPDAFQHASRD 636
 QY 600 LLSLVGDSKILATI 615
 DB 637 LLAAGIDDAAGVVAL 652

RESULT 9

B64172
 dse protein - Haemophilus influenzae (strain Rd KW20)
 C:Species: Haemophilus influenzae
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 24-Sep-1999
 C:Accession: B64172
 R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, B.F.; Kerlavage, A.
 J.; Gocayne, J.D.; Scott, J.; Shiley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
 D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.
 Science 269, 466-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; PMID:9530630; PMID:7542800
 A:Accession: B64172
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-625 <TIGR>
 A:Cross-references: GB:U22822; GB:L42023; NID:g1574265; PIDN:AAC23088.1; PID:g1574278; T
 A:Reference number: A59328
 A:Contents: annotation
 A:Gene: dse
 C:Superfamily: hypothetical protein C2814

Query Match 60.2%; Score 1937.5; DB 2; Length 625;
 Best Local Similarity 59.6%; Pred. No. 4.8e-128;
 Matches 368; Conservative 102; Mismatches 142; Indels 5; Gaps 3;

6 DYPPLKNIHTPADRALSKDLOOLADEVGYLTHTVTSISGHPAAGLGVETLVATLH 65
 DB 7 NYPLLSINSPEIDLRLNKDLPOLCQELRAYLLESVQSGLASGIGVETLVATLH 66
 QY 66 FNTFVDOLVMDVGHQAVPHKILTGKRRMPTITLGGVSAFPARDESEYDAFVGHSS 125
 DB 67 YKTPDOLVMDVGHQAVPHKILTGKRRMPTITLGGVSAFPARDESEYDAFVGHSS 126
 QY 126 ISALGKAIASOLGEBDKKVAIIIGDSITGGMAVEAMNHAGVNV--ANLTVLTIND 185
 DB 127 ISAGIYAVAEERNAERNAKRTVCVIGDGAITGAVFAELNHAALHTDMLVLTIND 186
 QY 186 PPGVGMANNLYTKVLSKFEYSVRESKCALAKMPSVWELAKTEETHVKMIV--PGTIF 243
 DB 187 ENVALNHNHARITSSGLYSTLRDSSKILDKVPIPNPKKTEETHVKMIV--PGTIF 246
 QY 244 ELGPNYFPIGDHVENMLVSTLENLKDITGPVFLHVTTKKGYAPAEKDLAHHG 303
 DB 247 ELGPNYFPIGDHVENMLVSTLENLKDITGPVFLHVTTKKGYAPAEKDLAHHG 306
 QY 304 DPTDFPKAAPSHPITVTFVGRWLCDMAAODERLLGTPAMREGSGLVFESQKF 363
 DB 307 DPTDFPKAAPSHPITVTFVGRWLCDMAAODERLLGTPAMREGSGLVFESQKF 364
 QY 364 PDVAIAEGHATVLAAGACGAKPVVAIYSTFLQSGYQLHVDVALQNLDMFLD 423
 DB 365 PDVAIAEGHATVLAAGACGAKPVVAIYSTFLQSGYQLHVDVALQNLDMFLD 424
 QY 424 VGPDPGPTHAGAFDYISYKRCIPNMLTMAPADENECROKLTTPQHHGPASVYPRGK 483

Db 425 VGADGATHQGAADISFMRGCIPTMTIPDSBENCGQMLYTGCGCKPAAYRPGNAVG 484
 Oy AIDPFLTLAIEGAEVREHSGRIATLAWGSMVTPAVEAGKQIGATVMMRFPVCKPDQALV 543
 Db 445 KLTLP-LEMLPICKSLIRKQKATILNFTLPSLSEISKLNATVMMRFPVCKPDQALV 543
 Oy 544 IELATADVFTVEENVIAGGASAINFLQAKVLMPCVNGILPDRFVGGSSREBLISL 603
 Db 544 NVLAQTHYLVTLERNAIOGGAGSAVAEVLNCSKSTALLQGLGPDVYFIPQATQCEALAD 603
 Oy 604 VGIDSKGILATEQPCA 620
 Db 604 LGDTKGIEBKLTNPIA 620

RESULT 10

B81978
 Probable 1-deoxyxylulose-5-phosphate synthase NMA0589 [imported] - Neisseria meningitidis
 C/Species: Neisseria meningitidis
 C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C/Accession: B81978
 R/Author: J. J. Achtman, M. J. James, K. D. Bentley, S. D. Churcher, C. Klee, S. R. Morel,
 R. Parkhill, J. J. Achtman, K. J. Leather, S. J. Mowle, S. J. Mungall, M. A. Rajandream,
 Nature 404, 502-506, 2000
 A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A/Reference number: A81775; MUID:20222556; PMID:10761919
 A/Accession: B81978
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-637 <PAR>
 A/Cross-references: GB:AL162753; GB:AL157959; NID:G7379120; PIDN:CA883880.1; PID:G737932
 A/Experimental source: serogroup A, strain Z2491
 C/Genetics:
 A/Gene: dks, NMA0589
 C/Superfamily: hypothetical protein C2814

Query Match 55.2%; Score 1776; DB 2; Length 637;
 Best Local Similarity 55.3%; Pred. No. 1.1e-116;
 Matches 347; Conservative 92; Mismatches 155; Indels 34; Gaps 5;

Oy 8 PLKNIHTPADIRALSKOQLADDEVKGLTHTVSISSGHPAAGLSTVELTVALHYVN 67
 Db 5 PLDLIDSPQDIRLDRKQPLRAGELRTLLSVGGQGHFASNLGAVELTVALHYVN 64
 Oy 68 TPVDQLVMDVGEQAYPHKILTGKERMPTITLGGSAFPARDESEYDAFGVGHSSSTIS 127
 Db 65 TPEDKLVMDVGEQAYPHKILTGKERMPTITLGGSAFPARDESEYDAFGVGHSSSTIS 124
 Oy 128 AALGMAIASQURGEDKXKVAIIGDSITGMAAYEAMNAGDVANLVLINDNMSISPP 187
 Db 125 AALGMAIASQURGEDKXKVAIIGDSITGMAAYEAMNAGDVANLVLINDNMSISPP 184
 Oy 188 VGAMNNYLTLYLSKPY---SSVRESKALAKMPSVMEIARKE-----EHYKG 234
 Db 185 VGALPKYLAISNVVDMGSLSTVAKOTGVLDKIPGAMERPAQKHKIKTLAEERHAKQ 244
 Oy 235 MIYVGTLEFEEIGFNTFGPIDGDEVMLVSTLENKDLTGSPFLHVVTKKGYAPAEKDP 294
 Db 245 SL---SLFENFGFRYTGPDGHNENLVLDLADSKSKGQQLLHVTIKKNGYGLAENDP 301
 Oy 295 LAYHGVPAFDPKDFLPSKAA-----PSHPYTEVFGMLCDMAQDERLLGIT 343
 Db 302 VKYHAVAN-----LPKESAQMPSEKPRPAKFTTYQVFGKMLCDRAADSLVAIT 354
 Oy 344 PAMRGGSLVGEFSQKFNRYFDVAIAEOHAVTLAAGACGAKPVVAIYSTFLQGYDOL 403
 Db 355 PAMRGGSLVGEFSQKFNRYFDVAIAEOHAVTLAAGACGAKPVVAIYSTFLQGYDOL 414
 Oy 404 IHDVALQNTLDMFLADBRAGVGPDPGTHAGFYDSYMRGCIPTMTIPADENECRQMLTT 463
 Db 415 VHDIALQNTLDMFLADBRAGVGPDPGTHAGFYDSYMRGCIPTMTIPADENECRQMLTT 474
 Oy 464 GFQHHGPAVSRYPRGKGPAAIDPTLTAIEGAEVREHSGRIATLAWGSMVTPAVEAGK 523

Db 475 CYQADPAAVRYRGGTGTGPVSDGMEVIEIGGIIIRREKTAFLAFSGMVAPALAVAG 534
 Oy 524 QLGATVMMRFPVCKPDQALVLEARTDVVVEENVIAGGASAINFLQAKVLMPCV 583
 Db 535 KLNATVADMRFPVCKPDQALVLEARTDVVVEENVIAGGASAINFLQAKVLMPCV 583
 Oy 584 NIGLPDRFVGGSSREBLISL 611
 Db 595 ILGAVDVTGKGDPKILDLGLSABAV 622

RESULT 11

D81034
 1-deoxyxylulose-5-phosphate synthase NMB1867 [imported] - Neisseria meningitidis (strain
 C/Species: Neisseria meningitidis
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C/Accession: D81034
 R/Author: H. J. Saunders, N. J. Heidelberg, J. J. Jeffries, A. C. Nelson, K. E. Eisele, J. A.
 Hickey, E. K. Haff, D. H. Salberg, S. L. White, O. Fleischmann, R. D. Dougherty, B. A.
 H. H. Qiu, H. Yamanevaru, J. Gill, J. Scarlato, V. Maignani, V. Pizze, M.
 Science 287, 1809-1815, 2000
 A/Authors: Grandi, G.; Sun, L.; Smith, H. O.; Fraser, C. M.; Noxon, E. R.; Rappuoli, R.; Ve
 A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MCS8.
 A/Reference number: A81000; MUID:20175755; PMID:10710307
 A/Accession: D81034
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-637 <TST>
 A/Cross-references: GB:AE002536; GB:AE002098; NID:G7227115; PIDN:AAE42201.1; PID:G722712
 A/Experimental source: serogroup B, strain MCS8
 C/Genetics:
 A/Gene: NMB1867
 C/Superfamily: hypothetical protein C2814

Query Match 55.2%; Score 1775; DB 2; Length 637;
 Best Local Similarity 54.8%; Pred. No. 1.2e-116;
 Matches 344; Conservative 97; Mismatches 153; Indels 34; Gaps 5;

Oy 8 PLKNIHTPADIRALSKOQLADDEVKGLTHTVSISSGHPAAGLSTVELTVALHYVN 67
 Db 5 PLDLIDSPQDIRLDRKQPLRAGELRTLLSVGGQGHFASNLGAVELTVALHYVN 64
 Oy 68 TPVDQLVMDVGEQAYPHKILTGKERMPTITLGGSAFPARDESEYDAFGVGHSSSTIS 127
 Db 65 TPEDKLVMDVGEQAYPHKILTGKERMPTITLGGSAFPARDESEYDAFGVGHSSSTIS 124
 Oy 128 AALGMAIASQURGEDKXKVAIIGDSITGMAAYEAMNAGDVANLVLINDNMSISPP 187
 Db 125 AALGMAIASQURGEDKXKVAIIGDSITGMAAYEAMNAGDVANLVLINDNMSISPP 184
 Oy 188 VGAMNNYLTLYLSKPY---SSVRESKALAKMPSVMEIARKE-----EHYKG 234
 Db 185 VGALPKYLAISNVVDMGSLSTVAKOTGVLDKIPGAMERPAQKHKIKTLAEERHAKQ 244
 Oy 235 MIYVGTLEFEEIGFNTFGPIDGDEVMLVSTLENKDLTGSPFLHVVTKKGYAPAEKDP 294
 Db 245 SL---SLFENFGFRYTGPDGHNENLVLDLADSKSKGQQLLHVTIKKNGYGLAENDP 301
 Oy 295 LAYHGVPAFDPKDFLPSKAA-----PSHPYTEVFGMLCDMAQDERLLGIT 343
 Db 302 VKYHAVAN-----LPKESAQMPSEKPRPAKFTTYQVFGKMLCDRAADSLVAIT 354
 Oy 344 PAMRGGSLVGEFSQKFNRYFDVAIAEOHAVTLAAGACGAKPVVAIYSTFLQGYDOL 403
 Db 355 PAMRGGSLVGEFSQKFNRYFDVAIAEOHAVTLAAGACGAKPVVAIYSTFLQGYDOL 414
 Oy 404 IHDVALQNTLDMFLADBRAGVGPDPGTHAGFYDSYMRGCIPTMTIPADENECRQMLTT 463
 Db 415 VHDIALQNTLDMFLADBRAGVGPDPGTHAGFYDSYMRGCIPTMTIPADENECRQMLTT 474
 Oy 464 GFQHHGPAVSRYPRGKGPAAIDPTLTAIEGAEVREHSGRIATLAWGSMVTPAVEAGK 523

Db 475 CYQADPAARVVRGTGTCGAPVSDGKMTVEIKGIIIRREGNTAFIAGSNVAPLAAVE 534
 QY 524 QLGATVVMNRFPKPPDQALVLEIARTHDVFTVEENVNIAAGAGSAINTEFLQAKYLMKVC 583
 Db 535 KLNATVADMRFPKPIDBELIVLARSHDRIVLLENAAGGAGAVLEIARHGICKEVLT 594
 QY 584 NIGLEDFRVEQSSREELSLVGLDSKGI 611
 Db 595 LLGVADVTYGRKDPKLLDDIGLSAAV 622

RESULT 12

das protein [imported] - Buchnera sp. (strain APS)

C.Species: Buchnera sp.
 C.Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C.Accession: A84984
 R.Shigenobu, S.; Matanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
 Nature 407, 81-86, 2000
 A.Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.
 A.Reference number: A84930; MUID:20445173; PMID:10993077
 A.Accession: A84984
 A.Status: Preliminary
 A.Molecule type: DNA
 A.Residues: 1-608 <STO>
 A.Cross-references: GB:AP000398; GSPDB:GN00144
 A.Experimental source: strain APS
 C.Genetics:
 A.Gene: dks; BU464
 C.Superfamily: hypothetical protein C2814

Query Match 51.8%; Score 1667; DB 2; Length 608;
 Best Local Similarity 52.8%; Pred. No. 4-4e-109;

Matches 325; Conservative 118; Mismatches 146; Indels 26; Gaps 7;

QY 7 YLLKNITPADIRALSKDQLOLADVEVGYLTFTVSISGHPAAGTVELITVALHYF 66
 Db 8 YPILSFANSVENIRLISVEQLPQLCFELREYLLDVVISISKHFAAGLGVETITVALHYV 67
 QY 67 NTPVDOLVMDVGHQAYPHKILNGRKEKMPITRTLGVSAPFARDESRYDAFGVSHSTSI 126
 Db 68 NTPEDNLMDGCHQAYPHKILNGRKEKINSIRKNGHSHFPERESSEDSISGVSHSTSI 127
 QY 127 SAALGMAIASQLRGEDKXWVAILIGDSITGCMAYEAMNAGDVANLVLINDNDMSIS 186
 Db 128 SAGLGMSIAAEKRGKRIKICIGGAMTAGMAFAINHHAGISQNLVILINDNDMSISR 187
 QY 187 PYGANNVITKYLSS-----KFSYVRESKKAALKMSVWEIAKTEEHVKMIVPGL 241
 Db 188 NGVALNKLH-KILRSVQNTQKRRKIRLNLKLFK-----DKRIQH--SISFNSI 236
 QY 242 FELLGFNYFGPIDGHDVEMLTLENKDLTGVPVLAHVYTKKQYAPAEKDPLAHHGVP 301
 Db 227 FSNLCCKVLGPGPDGNIFSIINTLKIKKKGTYLLHLYTKKQYLPALNPIKWHITIS 296
 QY 302 ADPPTKDFLPKAAEPHPHYTVEFGRLCDMAAOBERLLGITPARBESSGLVBSQKPPN 361
 Db 297 SBD-----SSVKSLSYSDVFGTWLCEIAAFPKKIATTPAACSSGMYKFSRLPN 348
 QY 362 RFVDVAIAEGEAVTLTAAGACQAGAPVAAYSTFGRGYDQILHVAIQNDMLFALDRA 421
 Db 349 QYFDVAIAEGEAVTLTAAGALISGYFVSYSTFGQRAYDQILHVAIQNKSVLFAVDRA 408
 QY 422 GLVPGDGPTHAGAFYSYRCIPNMLIMAPADENECRQMLTTFGQHH-GEASVRYRQKG 480
 Db 409 GIVGNDGGTHGQVFLAYLRCPGIVIMTPSNMENCROMLYGVYENKGPSPVRYPKYG 468
 QY 481 PCAALDPTLTALIEGAEVRRHSGRIALAMGSMTPAVEAGKCGATVVMNRFPKPPDQ 540
 Db 469 VQELLNP-MNRPIPKSLIKRQKIIAILNPSILHNAACAESKDAILVDRFPKPPDK 527
 QY 541 ALVLEIARTHDVFTVEENVNIAAGAGSAINTEFLQAKYLMKVCGLPDRVEQSSREEL 600

Db 528 SWILKLSQNKFPITLLEGVYISGAGSAAVEFIWVKILFPLVNLIGLPDPTIPQGTQEEI 587
 QY 601 LSLVGLDSKGIARTI 615
 Db 588 RHVYKLDSEGIYKQI 602

RESULT 13

A70376

conserved hypothetical protein aq_881 - Aquifex aeolicus

C.Species: Aquifex aeolicus
 C.Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 24-Sep-1999
 C.Accession: A70376

R.Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ore
 V.

Nature 392, 353-358, 1998

A.Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A.Reference number: A70300; MUID:98196666; PMID:9537320

A.Accession: A70376

A.Status: Preliminary; nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-628 <AGP>

A.Cross-references: GB:AE00712; MID:92983411; PIDN:AAC07004.1; PID:G2983421; GB:AE00065.
 A.Experimental source: strain VFS
 C.Genetics:
 A.Gene: aq_881
 C.Superfamily: hypothetical protein C2814

Query Match 50.4%; Score 1620.5; DB 2; Length 628;
 Best Local Similarity 51.9%; Pred. No. 8.6e-106;
 Matches 320; Conservative 105; Mismatches 174; Indels 17; Gaps 8;

QY 7 YLLKNITPADIRALSKDQLOLADVEVGYLTFTVSISGHPAAGTVELITVALHYF 66
 Db 5 YEIKDYKGPPIDKNYETLOKLAQEVVDYIINVTIRKNGHVPSPSLGVETITVALHYF 64
 QY 67 NTPVDOLVMDVGHQAYPHKILNGRKEKMPITRTLGVSAPFARDESRYDAFGVSHSTSI 126
 Db 65 NTPEDVLVMDIGHQYPMKILTDKXQFPLRQYKGISGFRRRESIYDAFGAGSHSTSI 124
 QY 127 SAALGMAIASQLRGE-DKXWVAILIGDSITGCMAYEAMNAGDVANLVLINDNDMSI 184
 Db 125 SAALGFRIGDILKEKEDYVLAIVGDALTRGMAVEALNNAHGRIPRFYIINDNMSI 184
 QY 185 SPVGMANNVITKYLSSKFSYVRESKKAALAN-PSWEIAKTEEHVKMIVPGLFE 243
 Db 185 SPNVGALSTYINRILISGFVQETROKIKNFLGHRGETPLRMKLTREFLKGLISPGVFE 244
 QY 244 ELGFNYFGPIDGHDVEMLTLENKDLTGVPVLAHVYTKKQYAPAEKDPLAHHGVP 303
 Db 245 ELGFNYFGPIDGHDILKLEDTLNNVKDIKGPVLLHVYTKKQYKPAENFVNMHGAAPY 304
 QY 304 D-PTKDFLPKAAEPHPHYTVEFGRLCDMAAOBERLLGITPARBESSGLVBSQKPPN 362
 Db 305 KVESGEIIRKSSP---PLTWSVFGKALVELABDEKIVALTIPARBESSGLVEFAKFPDR 361
 QY 363 YEDVAIAEGEAVTLTAAGACQAGAPVAAYSTFGRGYDQILHVAIQNDMLFALDRA 422
 Db 362 FFDVGAIBOHACTFPAAGIAEGAPVAAYSTFGQRAYDQILHVAIQNLPVTAIDPRAG 421
 QY 423 LVGPDGPTHAGAFYSYRCIPNMLIMAPADENECRQMLTTFGQHH-GEASVRYRQKG 482
 Db 422 LVGNDGPTHGVPFLSLRCVPRNVVCAKPDQELRLVLTGISGKFPALRYRGAAYG 481
 QY 483 AALDPT--LTALEIGAEVRRHSGRIALAMGSMTPAVEAGKCGATVVMNRFPKPPDQ 535
 Db 482 V--PTGSGFKKIEIGTVEELLEGDCVILAWGVPVYQALAAEKLVEGIRGVGNARFV 538
 QY 536 KPPDQALVLEIARTHDVFTVEENVNIAAGAGSAINTEFLQAKYLMKVCGLPDRVEQSS 595
 Db 539 KPNDBKRLDLNARYDPTFIVEDNTVVGSGVLEFFAABGIMKRYINLGVPRFIEHG 598
 QY 596 SREELSLVGLDSKGI 611

Db 599 KODILNIVGIDAGI 614

RESULT 14

AD3439
1-deoxyxylulose-5-phosphate synthase (EC 4.2.1.-) [imported] - *Brucella melitensis* (str. C) [Species: *Brucella melitensis*]
C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C/Accession: AD3439
R/RefSeq: V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, B.; Mazur, M.; Goldstein, E.; Salikov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A/Reference number: AD3439; PMID:1175668
A/Accession: AD3439
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-643 <KUR>
A/Cross-references: GB:AE008917; PIDN:AA152679.1; PID:gl7983505; GSPDB:GN00190
A/Experimental source: strain 16M
A/Genetics: BME11498
A/Map position: 1
A/Superfamily: hypothetical protein C2814
C/Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 49.5%; Score 1591.5; DB 2; Length 643;

Best Local Similarity 51.4%; Pred. No. 9, 6e-104; Mismatches 182; Indels 13; Gaps 7;

Matches 318; Conservative 106; Mismatches 182; Indels 13; Gaps 7;

8 PLKNIHTPADIALSKDQLOADVRGTLTTSISGHPAAGTVELTVALHYEN 67
7 PLIDKAPFDPRRLAEOLPLAEELRTELIDAVSTGGHAGAGVETLTAALHYEN 66
68 TPVDQVMDVGHQAYPHKILITGKEMPTIRITLQSVAPAPADESEYDAFGVHSSTIS 127
67 TPYDRITWVGHQAYPHKILITGRRDRIRITLQAGLSGFTKASESYDEFGAHSSTIS 126
128 AALGMAIASQLRGDKMVAIIIGDSITGMAVEANMHAGDVANLITLINDNIS 187
127 AGGMAVASLSEKKNVIAVIGDSMSKMAVEANMHAGDVANLITLINDNIS 186
188 VGAMNNYLTLYLSKFSYSSVRESKALAMP5-VNELAKTBEHYKGMIVPGTLFEEELG 246
187 TGAMSYLALVSGRTYSRBAKQVAKL.PKFLQDKARKSEYARAFPTGTLFEEELG 246
247 FNFPGIDHGDVEMVSTLENLKD-LTGPVFLAVTKKGGYAPAKDPLAHGVAPAF 305
247 FNFPGIDHGDVEMVSTLENLKD-LTGPVFLAVTKKGGYAPAKDPLAHGVAPAF 305
247 FNFPGIDHGDVEMVSTLENLKD-LTGPVFLAVTKKGGYAPAKDPLAHGVAPAF 305
247 FNFPGIDHGDVEMVSTLENLKD-LTGPVFLAVTKKGGYAPAKDPLAHGVAPAF 305
306 TNDLFKAPSPHPTTYEVRGRLCDMAADDERLIGITPAMEGSGLVEFSOKFPNRYD 365
307 ITGKQK-PKANPSTYKIFGTSLIEARHDKIVAVTAMPGTGLDFGEAFPRVFD 365
366 VALAEQHAATVLAGOACQAKPVAIYSTFLQGRYQOLIHVALQNLMDLPAIDRAGLV 425
366 VALAEQHAATVLAGOACQAKPVAIYSTFLQGRYQOLIHVALQNLMDLPAIDRAGLV 425
366 VALAEQHAATVLAGOACQAKPVAIYSTFLQGRYQOLIHVALQNLMDLPAIDRAGLV 425
366 VALAEQHAATVLAGOACQAKPVAIYSTFLQGRYQOLIHVALQNLMDLPAIDRAGLV 425
426 PDGPTAGAFDYSYMCIPNMLIMADENECQMLTTGFOH-HGASVRYPRGKPGAA 484
426 ADGPTAGAFDYSYMCIPNMLIMADENECQMLTTGFOH-HGASVRYPRGKPGAA 484
485 IDPTLALBETGKAEVHHSRIAILANGSVTPAVEAGQOLG-----TVNMRPYKPD 539
485 IDPTLALBETGKAEVHHSRIAILANGSVTPAVEAGQOLG-----TVNMRPYKPD 539
485 IDPTLALBETGKAEVHHSRIAILANGSVTPAVEAGQOLG-----TVNMRPYKPD 539
485 IDPTLALBETGKAEVHHSRIAILANGSVTPAVEAGQOLG-----TVNMRPYKPD 539
540 QALVLEARTHDVFTVVEENVVAGAGSAINFLQAKVLT---MPVCNIGLPDRFEGGS 596
540 QALVLEARTHDVFTVVEENVVAGAGSAINFLQAKVLT---MPVCNIGLPDRFEGGS 596
540 QALVLEARTHDVFTVVEENVVAGAGSAINFLQAKVLT---MPVCNIGLPDRFEGGS 596
540 QALVLEARTHDVFTVVEENVVAGAGSAINFLQAKVLT---MPVCNIGLPDRFEGGS 596
546 HDLIRLAEHEVTVVVEBAV-GGFGSHVLOFLADGLDGRFKVRLALTLPIYDGRK 604
546 HDLIRLAEHEVTVVVEBAV-GGFGSHVLOFLADGLDGRFKVRLALTLPIYDGRK 604
546 HDLIRLAEHEVTVVVEBAV-GGFGSHVLOFLADGLDGRFKVRLALTLPIYDGRK 604
546 HDLIRLAEHEVTVVVEBAV-GGFGSHVLOFLADGLDGRFKVRLALTLPIYDGRK 604
597 REELSLVGLDSKGLIATI 615
597 REELSLVGLDSKGLIATI 615
597 REELSLVGLDSKGLIATI 615
597 REELSLVGLDSKGLIATI 615

Db 605 PDAMYARGLDRGTIVRTV 623

RESULT 15

C97505
1-deoxyxylulose-5-phosphate synthase [imported] - *Caulobacter crescentus*
C/Species: *Caulobacter crescentus*
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C/Accession: C97505
R/RefSeq: M.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Bisen, J.; Heidelberg, J. B.; Land, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gilm, M.L.; Haft, D.H.; Kolon
n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of *Caulobacter crescentus*.
A/Reference number: A67249; MUID:21173696; PMID:11259667
A/Accession: C97505
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-640 <STO>
A/Cross-references: GB:AE005673; NID:gl3423547; PIDN:AAK24039.1; GSPDB:GN00148
A/Genetics: CC2068
A/Superfamily: hypothetical protein C2814

Query Match 47.4%; Score 1523.5; DB 2; Length 640;

Best Local Similarity 50.7%; Pred. No. 5, 7e-99; Mismatches 196; Indels 13; Gaps 7;

Matches 315; Conservative 97; Mismatches 196; Indels 13; Gaps 7;

3 LTTDPLKNIHTPADIALSKDQLOADVRGTLTTSISGHPAAGTVELTVAL 62
1 MSSKPLIDTASPADTGLSLAEIKOLAAEVRAETIDAVSTGGHAGAGVETLTAAL 60
63 HYVNTPVQDQVMDVGHQAYPHKILITGKEMPTIRITLQSVAPAPADESEYDAFGVHS 122
61 HNVETPDVIVMDVGHQAYPHKILITGRRDRIRITLQAGLSGFTKASESYDEFGAHS 120
123 STSISALGMAIASQLRGDKMVAIIIGDSITGMAVEANMHAGDVANLITLINDNM 182
121 ATSISALGMAIASQLRGDKMVAIIIGDSITGMAVEANMHAGDVANLITLINDNM 180
183 SISFPVGMNNYLTLYLSKFSYSSVRESKALAMP5-VNELAKTBEHYKGMIVPGTL 241
181 SIAPFVGMNNYLTLYLSKFSYSSVRESKALAMP5-VNELAKTBEHYKGMIVPGTL 240
242 FEEELGFTYFGIDHGDVEMVSTLENLKD-LTGPVFLAVTKKGGYAPAKDPLAHGV 300
241 FEEELGFTYFGIDHGDVEMVSTLENLKD-LTGPVFLAVTKKGGYAPAKDPLAHGV 300
301 PAEDPTKDFLKAAPSPHPTTYEVRGRLCDMAADDERLIGITPAMEGSGLVEFSOKFP 360
301 VEDPTKDFLKAAPSPHPTTYEVRGRLCDMAADDERLIGITPAMEGSGLVEFSOKFP 360
301 VEDPTKDFLKAAPSPHPTTYEVRGRLCDMAADDERLIGITPAMEGSGLVEFSOKFP 360
301 VEDPTKDFLKAAPSPHPTTYEVRGRLCDMAADDERLIGITPAMEGSGLVEFSOKFP 360
361 NRYFPAIAEQAATVLAGOACQAKPVAIYSTFLQGRYQOLIHVALQNLMDLPAIDR 420
360 ERTFPAIAEQAATVLAGOACQAKPVAIYSTFLQGRYQOLIHVALQNLMDLPAIDR 419
421 AGLVGDPGPTAGAFDYSYMCIPNMLIMADENECQMLTTGFOH-HGASVRYPRGKPG 479
420 AGLVGDPGPTAGAFDYSYMCIPNMLIMADENECQMLTTGFOH-HGASVRYPRGKPG 479
480 GPGAIIDPTLALBETGKAEVHHSRIAILANGSVTPAVEAGQOLG-----AGKOLGATVNNRF 534
480 GPGAIIDPTLALBETGKAEVHHSRIAILANGSVTPAVEAGQOLG-----AGKOLGATVNNRF 534
480 GPGAIIDPTLALBETGKAEVHHSRIAILANGSVTPAVEAGQOLG-----AGKOLGATVNNRF 534
480 GPGAIIDPTLALBETGKAEVHHSRIAILANGSVTPAVEAGQOLG-----AGKOLGATVNNRF 534
540 AKPLDLDLLELAEHEVTVVVEBAV-GGFGSHVLOFLADGLDGRFKVRLALTLPIYD 598
540 AKPLDLDLLELAEHEVTVVVEBAV-GGFGSHVLOFLADGLDGRFKVRLALTLPIYD 598
540 AKPLDLDLLELAEHEVTVVVEBAV-GGFGSHVLOFLADGLDGRFKVRLALTLPIYD 598
540 AKPLDLDLLELAEHEVTVVVEBAV-GGFGSHVLOFLADGLDGRFKVRLALTLPIYD 598
592 VEGSREELSLVGLDSKGLIATI 612
592 VEGSREELSLVGLDSKGLIATI 612
592 VEGSREELSLVGLDSKGLIATI 612
592 VEGSREELSLVGLDSKGLIATI 612
599 QDQKRPDMYTAQAGIDAGIL 619
599 QDQKRPDMYTAQAGIDAGIL 619
599 QDQKRPDMYTAQAGIDAGIL 619
599 QDQKRPDMYTAQAGIDAGIL 619

Thu Jan 29 17:41:29 2004

us-09-941-947a-6.rpt

Page 9

Search completed: January 29, 2004, 15:55:29
Job time : 20.2424 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 11.6752 Seconds

(without alignment)
2497.314 Million cell updates/sec

Title: US-09-941-947a-6

Sequence: 1 MCLTIDYPLKKNITPADIR.....LSLVGLDSKGIATTEQPCA 620

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2114	65.7	621	1	DXS_VIBVU
2	2097.5	65.2	626	1	DXS_VIBCH
3	2069.5	64.4	627	1	DXS_VIBCH
4	2042	63.5	622	1	DXS_SHEON
5	2015	62.7	619	1	DXS_YERPB
6	2014.5	62.6	619	1	DXS_SALTI
7	2013.5	62.6	619	1	DXS_SALTY
8	2008.5	62.5	636	1	DXS_RALSO
9	2000.5	62.2	619	1	DXS_ECOLI
10	1999.5	62.2	619	1	DXS_ECOLI
11	1995.5	61.9	619	1	DXS_ECOLI
12	1989.5	61.9	638	1	DXS_XANCP
13	1986.5	61.8	638	1	DXS_XANCP
14	1969	61.2	614	1	DXS_PASMU
15	1951	60.7	635	1	DXS_XYFPA
16	1937.5	60.2	635	1	DXS_HATIN
17	1795	55.8	626	1	DXS_WIGBR
18	1776	55.2	637	1	DXS_NEIMA
19	1775	55.2	637	1	DXS_NEIMA
20	1667	51.8	608	1	DXS_BUCAI
21	1629.5	50.7	585	1	DXS_BUCAP
22	1620.5	50.4	628	1	DXS_AQUAE
23	1592.5	49.5	643	1	DXS_BRUSU
24	1591.5	49.5	643	1	DXS_BRUSU
25	1543.5	48.0	637	1	DXS_RHIO
26	1523.5	47.4	640	1	DXS_CAUCR
27	1516.5	47.2	622	1	DXS_THETN
28	1499.5	46.6	639	1	DXS_AGRTS
29	1498.5	46.6	639	1	DXS_AGRTS
30	1460.5	45.4	635	1	DXS_RHIME
31	1459	45.4	635	1	DXS_RHIME
32	1455	45.2	628	1	DXS_BACD
33	1442	44.8	640	1	DXS_SYMY3

ALIGNMENTS

RESULT 1	ID	DXS_VIBVU	STANDARD:	PRT:	621 AA.
AC	Q8DFA3				
DT	15-SEP-2003 (rel. 42, Last sequence update)				
DT	15-SEP-2003 (rel. 42, Last sequence update)				
DT	15-SEP-2003 (rel. 42, Last sequence update)				
DE	1-deoxy-D-xylose-5-phosphate synthase (EC 2.2.1.7) (1-deoxy-D-xylose-5-phosphate synthase) (DXP synthase) (DXPS).				
DE	DXS OR V10315.				
OS	Vibrio vulnificus.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrrio.				
OX	NCBI_TaxID=672;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-CMCP6;				
RA	Rhee U.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H., Choy H.E.;				
RT	*Complete genome sequence of Vibrio vulnificus CMCP6."				
RL	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.				
CC	- FUNCTION: Catalyzes the acylol condensation reaction between C atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield 1-deoxy-D-xylose-5-phosphate (DXP) (By similarity).				
CC	- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-deoxy-D-xylose-5-phosphate + CO(2).				
CC	- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By similarity).				
CC	- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.				
CC	- SUBUNIT: Homodimer (By similarity).				
CC	- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL: AB016798; AAC08845.1; -				
DR	HAMAP: MF_00315; -; 1.				
DR	InterPro: IPR005477; Dxs.				
DR	InterPro: IPR005476; Transketolase_C.				
DR	InterPro: IPR005475; Transketolase_N.				
DR	InterPro: IPR005474; Transketolase_C.				
DR	Pfam: PF02780; transketolase_C; 1.				
DR	Pfam: PF02779; transket pyr; 1.				
DR	TIGRfam: TIGR00204; dxe; 1.				
DR	PROSITE: PS00801; TRANSKETOLASE_1; 1.				
DR	PROSITE: PS00802; TRANSKETOLASE_2; 1.				
KW	Transferrase; Flavoprotein; Thiamine pyrophosphate.				
KW	Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.				
SQ	SEQUENCE 621 AA; 67959 MW; 69DABA8B22B8D4B CRC64;				

Query Match 65.7%; Score 2114; DB 1; Length 621;
 Best Local Similarity 63.1%; Pred. No. 2,9e-141;
 Matches 392; Conservative 102; Mismatches 123; Indels 4; Gaps 2;

3 LTLD---YPLKNHTPADIRALSKDQLADVRGYTHVTSISGGHFAAGTVELT 59
 1 MTLDISKPTLALANTDELRSLRKEVLPKICDELRTYLLNSVSSGHLASGLTVELT 60
 60 VALHYVNTPVDOVLWDVGHQAYPHKILTRKEREMPTIRTLGVSAFPAARDESEYDAFV 119
 61 VALHYVNTPVDOVLWDVGHQAYPHKILTRKEREMPTIRTLGVSAFPAARDESEYDLSV 120
 120 GHSSTISALGMAIAQOLGEGDKKVAIIGDGSITGGAAYANMHADGVANLVTIIND 179
 121 GHSSTISALGMAIAQOLGEGDKKVAIIGDGSITGGAAYANMHADGVANLVTIIND 180
 180 NMSISPPVGMANNVLTYSKFSYSSVRESKALAKPVSWEELARKTEBEVKMIVPG 239
 181 NMSISENVGALNNHARVLSGNLYTSIREGKYLGSVPKELVTRTEHLKGMVVP 240
 240 TLFEELEFNYPIDGHDVEMVSTLENKDTGPFVFLAVTKKGGYAPAEKOPLAYHG 299
 241 TLFEELEFNYPIDGHDVEMVSTLENKDTGPFVFLAVTKKGGYAPAEKOPLAYHG 300
 300 VPAPDPTKDFPKAPSPHYTEVFRWLCDMAODERLGITPAMEGSGLVFEFSQKF 359
 301 VKPDPANHSIPKSS-GGKPSFSTNFGDPLCDMAODPTLMAITPAMEGSGLVFEFSQKF 359
 360 PNRYFDVALAEQAVTLAAGQACQAKPVVAISTFLOGYDQILHVALONLMLPALD 419
 360 PNRYFDVALAEQAVTLAAGQACQAKPVVAISTFLOGYDQILHVALONLMLPALD 419
 420 RAGLVGDPDPTAGAFDVSVCIPNMLIMADNRECKOMLTGPHHGSATVYPPKX 479
 420 RAGLVGDPDPTAGAFDVSVCIPNMLIMADNRECKOMLTGPHHGSATVYPPKX 479
 480 GPGAAIDPTLALTEIGKAEVRRHRSRIALLANGSWTPVAEAKOIGATVVMREYKPEFD 539
 480 GPGTPIESFTALTEIGKRLVYQGEKVALISFTGFLANALEAELNMTVADMRVPEFD 539
 540 QALVLELATHTVPTVVEENVLAGAGSAINTFLAQKILMVVCHGIDPRVVEGSSBE 599
 540 EALLSQALNEHVDVLTTEENALAGAGAGVIEFPMQEKIMKPVNLIDPKFIHQGTBE 599
 600 LLSLVGDSKGIATIEFCFA 620
 600 LHEELGIDAKGIGQAIRHYLA 620

RESULT 2
 DKS_VIBCH STANDARD; PRT; 626 AA.
 AC 09KTL3;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.1.7) (1-
 deoxyxylose-5-phosphate synthase) (DXP synthase) (DXPS).
 DKS OR VC0889.
 GN Vibrio cholerae.
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 CC NCBI_TaxID=666;
 RN 11;
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 Dodson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Unayan L.A.,
 Gill S.R., Nelson K.E., Read T.D., Tellein H., Richardson D.,
 Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Diracol I., Sellers P.,
 McDonald L., Ueberack T., Fleischmann R.D., Nieman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae.";
 RL Nature 406:477-483(2000).
 CC -1- FUNCTION: Catalyzes the acylotin condensation reaction between C
 atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
 1-deoxy-D-xylose-5-phosphate (DXP) (by similarity).
 CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-
 deoxy-D-xylose-5-phosphate + CO(2).
 CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (by
 similarity).
 CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol, first
 step.
 CC -1- SUBUNIT: Homodimer (by similarity).
 CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use. By non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AE004173; AAP94051.1; -
 PIR; H82266; H82266.
 DR TIGR; VC0889; -
 DR HAMAP; MF_00315; -; 1.
 DR InterPro; IPR005477; Dxs.
 DR InterPro; IPR005476; Transketolase C.
 DR InterPro; IPR005475; Transketolase CR.
 DR InterPro; IPR005474; Transketolase N.
 DR Pfam; PF02778; transket_pyr; 1.
 DR Pfam; PF02780; transketolase_C; 1.
 DR TIGRFAMs; TIGR00204; dxs; 1.
 DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
 DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
 DR Transfaser; Flavoprotein; Thiamine pyrophosphate;
 KW Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
 SQ SEQUENCE 626 AA; 68346 MW; 5838F99BE851C9 CRC64;

Query Match 65.2%; Score 2097.5; DB 1; Length 626;
 Best Local Similarity 62.5%; Pred. No. 4,2e-140;
 Matches 391; Conservative 109; Mismatches 117; Indels 9; Gaps 3;

3 LTLD---YPLKNHTPADIRALSKDQLADVRGYTHVTSISGGHFAAGTVELT 59
 1 MTLDISKPTLALANTDELRSLRKEVLPKICDELRTYLLNSVSSGHLASGLTVELT 60
 60 VALHYVNTPVDOVLWDVGHQAYPHKILTRKEREMPTIRTLGVSAFPAARDESEYDAFV 119
 61 VALHYVNTPVDOVLWDVGHQAYPHKILTRKEREMPTIRTLGVSAFPAARDESEYDLSV 120
 120 GHSSTISALGMAIAQOLGEGDKKVAIIGDGSITGGAAYANMHADGVANLVTIIND 179
 121 GHSSTISALGMAIAQOLGEGDKKVAIIGDGSITGGAAYANMHADGVANLVTIIND 180
 180 NMSISPPVGMANNVLTYSKFSYSSVRESKALAKPVSWEELARKTEBEVKMIVPG 239
 181 NMSISENVGALNNHARVLSGNLYTSIREGKYLGSVPKELVTRTEHLKGMVVP 240
 240 TLFEELEFNYPIDGHDVEMVSTLENKDTGPFVFLAVTKKGGYAPAEKOPLAYHG 299
 241 TLFEELEFNYPIDGHDVEMVSTLENKDTGPFVFLAVTKKGGYAPAEKOPLAYHG 300
 300 VPAPDPTKDFPKAPSPHYTEVFRWLCDMAODERLGITPAMEGSGLVFEFSQKF 359
 301 VKPDPANHSIPKSS-NYKTPFSKIFGDFLCDMAODPTLMAITPAMEGSGLVFEFSQKF 359
 360 PNRYFDVALAEQAVTLAAGQACQAKPVVAISTFLOGYDQILHVALONLMLPALD 419

DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.1.7) (1-
 deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS).
 GN DXP OR S01525.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NR-1.
 RX MEDLINE=22297666; PubMed=12368913;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Mehta B., Clayton R.A.,
 RA Meyer T., Tappin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
 RA Meller J., Kouri H., Gill J., Utterback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
 RT Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT *Shewanella oneidensis*.
 RL Nat. Biotechnol. 20:1118-1123(2002).
 CC - FUNCTION: Catalyzes the acylol condensation reaction between C
 CC atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
 CC 1-deoxy-D-xylose-5-phosphate (DXP) (By similarity).
 CC - CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-
 CC deoxy-D-xylose-5-phosphate + CO(2).
 CC - COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By
 CC similarity).
 CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.
 CC - PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
 CC step.
 CC - SUBUNIT: Homodimer (By similarity).
 CC - SIMILARITY: Belongs to the transketolase family. DXPS subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AE015598; AAN54586.1; .
 CC TIGR: S01525; .
 DR HAMAP: MF_00315; . 1.
 DR InterPro: IPR005477; Dxs.
 DR InterPro: IPR005476; Transketolase_C.
 DR InterPro: IPR005475; Transketolase_CR.
 DR InterPro: IPR005474; Transketolase_N.
 DR Pfam: PF02780; transketolase_C.1.
 DR Pfam: PF02779; transket_pyr.1.
 DR TIGRfam: TIGR00264; dxs.1.
 DR PROSITE: PS00801; TRANSKETOLASE_1; 1.
 DR PROSITE: PS00802; TRANSKETOLASE_2; 1.
 DR Transferrase; Flavoprotein; Thiamine pyrophosphate;
 KM Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
 SQ SEQUENCE 622 AA; 68083 MW; BE681C01BEA5EFB4 CRG4;
 Query Match 63.5%; Score 2042; DB 1; Length 622;
 Best Local Similarity 62.0%; Pred. No. 3,4e-136;
 Matches 383; Conservative 105; Mismatches 126; Indels 4; Gaps 3;
 QY 5 TDYPLKNTFPADIALSKDQLADEVRYGLTFTVSGHSPAGLGYELVTAALY 64
 DB 6 SCPPVLAQNTPEMLKQLPOLALLPQLADELREFLSVGSSCHFAISGLTVELTALY 65
 QY 65 VPTPVLDQVLDVDEGQAYPKKILTKRKMPTIRTKGVSAPAFADSESYDAFGVSHST 124
 DB 66 VYNTPEDRILWDVGHQAYPKKILTKRDMETIRKQKGLHPFPWRESEYDTFVSHSGT 125

QY 125 SISAAAGAAISQLEGEDEKXWVAIIIGDSITGMAVEAMNAGDVANALVINDNMSI 184
 DB 126 SISAAAGAAVAAEKQAKRVAIVIGDAMTGMGVFEMNAGDILNDMLVINDNMSI 185
 QY 185 SPPVQANNNYILTKVLSKFTSSVRESKALAKPSVWELARKTEHKKMIVETLTREE 244
 DB 186 SNNVALNNHNLQAOLMSGFPTTIRSSKRVKQGVIEKMKRTEELKGMVVPETLFE 245
 QY 245 IGFNYFGPIDGDHVMYLTSTLENLKDITGPVFLHVTYKKGKYAPAEKDPYAHGVPAPD 304
 DB 246 IGFNYIGPIDHDVVALVETILNNMNLKGPVLTMTTKGSGYBAEKDPIGMHVPAPD 305
 QY 305 PTKDPLPDAAPSPH-PITTEYFGKMLCMAAODERLGITPAMRSGQLVEPSQFENRY 363
 DB 306 PS--LEKKPATPKPGLPTPSQVFGKMLCDIACODEKVLGITPAMRSGQVVEPSQFENRY 363
 QY 364 FVVAIAEGHATVLAAGACOGAKPVVAIYSTPELORGXQLHDVALQMLDMLFALDRAGL 423
 DB 364 FVVAIAEGHATVLAAGACOGAKPVVAIYSTPELORGXQLHDVALQMLDMLFALDRAGL 423
 QY 424 VGPDPPTAGAFDVSVMCIENMLIMADENECRQMLTTGPQHH-GPASVRYPRKGP 482
 DB 424 VQADGPTQGAFFDLSFRCIPMWTMAPSDNECQQLYTCYDAGPSAVRYPRGSATG 483
 QY 483 AAIDPTLALILGKAEVHHGSRITAILAWGSMVTPAVAGKQLGATVNMFPVXEPQAL 542
 DB 484 ANQVEMATLPLFGKGVIRLGRILNLFITLALNATRAISLDTNVDFVXEPQAL 543
 QY 543 VLELARTHDVPTVEENVIAGAGSANTFLQAKVLMVPCVNIQIPDSFVQGSRELLS 602
 DB 544 VEMAGTIDVAVTVENNAIMGAGSGVLELLQKMPKPVLTQIGLPBFIKHSDEVIH 603
 QY 603 LVGDSKGLIATIEPFA 620
 DB 604 DLQIDAEKMLQINAYLA 621
 RESULT 5
 ID DXS_YERPE STANDARD; PRT; 619 AA.
 AC 08ZC45;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.1.7) (1-
 DE deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS).
 GN DXS OR YP03177 OR Y1008.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parthill U., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Felwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
 RT Genome sequence of *Yersinia pestis*, the causative agent of plague.*;
 RL Nature 413:523-527(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIMS / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Mason J.S., Blattner F.R.,

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AL627266; CAD08878.1; -

DR EMBL; AB016842; AAO70031.1; -

DR HAMAP; MF_00315; -; 1.

DR InterPro; IPR005477; Dxs.

DR InterPro; IPR005476; Transketolase C.

DR InterPro; IPR005475; Transketolase CR.

DR InterPro; IPR005474; Transketolase_N.

DR Pfam; PF02779; transket_pyr; 1.

DR TIGRPFAM; TIGR00204; dxs; 1.

DR PROSITE; PS00801; TRANSKETOLASE_1; 1.

DR PROSITE; PS00802; TRANSKETOLASE_2; 1.

KW Transferrase; flavoprotein; thiamine pyrophosphate;

KW Isoprene biosynthesis; thiamine biosynthesis; Complete proteome.

FT INIT MET 0

FT SEQUENCE 619 AA; 67364 MW; E9DYC61243C29DB7 CRC64;

Query Match 62.6%; Score 2014.5; DB 1; Length 619;

Best local Similarity 61.0%; Pred. No. 2,9e-134;

Matches 375; Conservative 106; Mismatches 131; Indels 3; Gaps 3;

QY 7 YPLKNIHTPADIRALSQDQLADVERGYLTHTVVISGHPAAGLTVELTVLHYV 66

DB 7 YPLTALVDSQTEIRLLPKESLPCLCELRRLYLDSSRSRSGHPSAGLTVELTVLHYV 66

QY 67 NTFVDQLVWDVGHQAPPHKILTGKREKMPITRTLGVSAPFPADESEYAFVGHSSSTSI 126

DB 67 NTFPDQLVWDVGHQAPPHKILTGKREKMPITRTLGVSAPFPADESEYAFVGHSSSTSI 126

QY 127 SAALGNALASQLRGDKKXVAITIGDSITGMAVYANFAGVANNALVILNDKMSISP 186

DB 127 SAQIGIAVAABKCKGKORRTWCVIGDAITAGMAFEMNHGDIRPMLVILNDKMSISP 186

QY 187 PVGANNNTLTKYLSKRYSSVRSSEKALAKMPSVWEIARKTEHVKMIVPGLFEEIG 246

DB 187 NVGALNNHARLISGKLYSLRGGKVFSGVPITELKTEHVKMIVPGLFEEIG 246

QY 247 FNFEGPIDGHVEMLVSTLENLKDGLGPELVAVTTKKKGVAPEAKOPLAHGVAPDPT 306

DB 247 FNVIGPDGHVEMLVSTLENLKDGLGPELVAVTTKKKGVAPEAKOPITFHAVPKPDPS 306

QY 307 KDELPKAAEPHPHYTVTFGRMLCKMAOBERLLGTPARBSGCVESOKENNYFVY 366

DB 307 SGCLPKSS - GGLPKYSKIPEDMPCETAKUSKLMALTPARESGGWESRKEPDYFVY 365

QY 367 AIAEQHVAITLAAQACGAKFVVAIVSTFQRGYDQLIHDVALONTDMLFALDRAGLVGP 426

DB 366 AIAEQHVAITLAAQACGAKFVVAIVSTFQRAVDQVTHVALQKLPVMAIDRAGIVGA 425

QY 427 DGFTHAGAFDYSYNECIPTNMLIAPADENSCROMLTTFGQHN - GPASVRRPKGPRAL 485

DB 426 DGFTHAGAFDYSYNECIPTNMLIAPADENSCROMLTTFGQHN - GPASVRRPKGPRAL 485

QY 486 DPTTALIEIGAVRHSGRIALILAMGSMYTPAVEAKGQATVYNNRFPYEPDQALVLE 545

DB 486 TP - LEKPIRGKGVKRGKELAILNPGTLMPEAKVABALNATLVDNRFPYEPDQALVLE 544

QY 546 LARTHDVFTVEENVVIAAGASAINTLQOKTLPVNCIGLPRFPKQSGRELLISVG 605

DB 545 MAAGHDLAVLLEENALIMGASGVNEVLMARKEPVNLINGLPDFTIPQIGEEARDELG 604

QY 606 LPSKGLATITBQCA 620

DB 605 LPAAGIEKIKANMLA 619

RESULT 7

DXS - SALTY

ID - DKS SALTY

AC Q88RD1; STANDARD; PRT; 619 AA.

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE 1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.1.7) (1-deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPs).

GN DKS OR STM0422.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI_Taxid=602;

RA [1]

RA SEQUENCE FROM N.A.

RC STEBIN-LT2 / SGSC1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.M., Latreille P., Courtney L., Portolillo S., Ali J., Dante M., Du P., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;

RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";

RL Nature 413:852-856(2001).

CC - FUNCTION: Catalyzes the acylotin condensation reaction between C atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield 1-deoxy-D-xylose-5-phosphate (DXP) (By similarity).

CC - CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-deoxy-D-xylose 5-phosphate + CO2.

CC - COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By similarity).

CC - PATHWAY: Nonnevalonate terpenoid biosynthesis pathway; first step.

CC - PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first step.

CC - SUBUNIT: Homodimer (By similarity).

CC - SIMILARITY: Belongs to the transketolase family. DXPs subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE008715; AAL19376.1; -

DR StycGene; SG27272; dxs.

DR HAMAP; MF_00315; -; 1.

DR InterPro; IPR005477; Dxs.

DR InterPro; IPR005476; Transketolase C.

DR InterPro; IPR005475; Transketolase CR.

DR InterPro; IPR005474; Transketolase_N.

DR Pfam; PF02779; transket_pyr; 1.

DR Pfam; PF02780; transketolase_C; 1.

DR TIGRPFAM; TIGR00204; dxs; 1.

DR PROSITE; PS00801; TRANSKETOLASE_1; 1.

DR PROSITE; PS00802; TRANSKETOLASE_2; 1.

KW Transferrase; flavoprotein; thiamine pyrophosphate;

KW Isoprene biosynthesis; thiamine biosynthesis; Complete proteome.

FT INIT MET 0

FT SEQUENCE 619 AA; 67336 MW; FB3525B91DB6F346 CRC64;

Query Match 62.6%; Score 2013.5; DB 1; Length 619;

Best local Similarity 61.0%; Pred. No. 3,4e-134;

Matches 375; Conservative 106; Mismatches 131; Indels 3; Gaps 3;

QY 7 YPLKNIHTPADIRALSQDQLADVERGYLTHTVVISGHPAAGLTVELTVLHYV 66

DB 7 YPLTALVDSQTEIRLLPKESLPCLCELRRLYLDSSRSRSGHPSAGLTVELTVLHYV 66

QY 67 NTFVDQLVWDVGHQAPPHKILTGKREKMPITRTLGVSAPFPADESEYAFVGHSSSTSI 126

DB 67 NTFPDQLVWDVGHQAPPHKILTGKREKMPITRTLGVSAPFPADESEYAFVGHSSSTSI 126

QY 127 SAALGNALASQLRGDKKXVAITIGDSITGMAVYANFAGVANNALVILNDKMSISP 186

```

DB 127 SAGIGIAVAABKEGDRRTVCITGGCAITAGACAFEMAHAGDIRDMLVILINDMSISB 186
QY 187 PYGANNVLTLYLSKFTSSVAREESKALAKRPSWELAKTHEEVKMINVDTLFEELG 246
DB 187 NVGALNNHIAQLLSGLVSSLSREGEKQVSSGVPIKEILKRTTEHIIKQVAVPGLTLEELG 246
QY 247 FUYFPIIDHADVEMVSTLENKDLNGPVYLVHVKKGKGYAPAKEDPLAYHGPADPPT 306
DB 247 FUYIGPVCHDVMGLISTLKNNRDLKGPFLIMTKKGGYEPKADDPITTEPAVYKPPDS 306
QY 307 KDFLPYAPSPHPTTEVFGRMLCDMAADERLLGITTPANREGSGIVFESQCFENRYDY 366
DB 307 SCGLPKRSS-GGLPGYSKIFGDMICETAAADSKLMAITPMBGSGWVEFSRFTPPRYDY 365
QY 367 ATAEGHATVLAAGCQAGKAPVAIYSTLOSQYOLIHVALONLMDLFAIDRAGLVGP 426
DB 366 ATAEGHATVLAAGLAIQKVPVVAIYSTLOSQYOLIHVALONLMDLFAIDRAGLVGP 425
QY 427 DGPTHAGAFDYSYMRCPINMLIMAPADENECRQMLTTFQHH-GRASVYRPGKGPAAI 485
DB 426 DQCTHOGAFDLTYLRICIPMVINTPSEDENECRQMLFTGYHVDGPTAVAYPRGNAQVYL 485
QY 486 DDTLTALTEIGKAEVRRHRSRIALLMGSWTFAVEAGKQAGTVVMRPFVPGDALVLE 545
DB 486 TP-LEKLPYIGKGLVKKHSEKLAINEGTLMPBAKVAEALNLDVDRKPLDITLIS 544
QY 546 LARTHDVFTVEENVTAGAGSAINFTLOAKVLMPCNIGLDPREVEGSRBELLSVWG 605
DB 545 MAAGHVALVTLLENALMGAGSGVNEVMAHRRKPVAVNLITGLPDRFIPQGTGBARAELG 604
QY 606 LDKGILATIEQCA 620
DB 605 LDAAGIEAKIKMIA 619

RESULT 8
DVS_RALSO STANDARD; PRT; 636 AA.
ID DVS_RALSO
AC 08XX95;
DC 28-FEB-2003 (Rel. 41, Created)
DC 28-FEB-2003 (Rel. 41, Last sequence update)
DC 15-SEP-2003 (Rel. 42, Last annotation update)
DB 1-deoxy-D-xylose 5-phosphate synthase (BC 2.2.1.7) (1-
  deoxyxylose-5-phosphate synthase) (DXP synthase) (DXS).
  DXS OR KSC221 OR RS01378.
  GN Ralstonia solanacearum (pseudomonas solanacearum).
  OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
  OC Burkholderiaceae; Ralstonia.
  OX NCBI_TaxID=305;
  RN (1)
  RP SEQUENCE FROM N.A.
  RC STRAIN=GM11000;
  RC MEDLINE=21681879; PubMed=11823852;
  RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
  RA Arlat M., Billault A., Broctier P., Camus J.C., Catolico L.,
  RA Gaspin C., Choisme N., Claudel-Renaud C., Cunac S., Demange N.,
  RA Chaudin C., Lave M., Moisan A., Robert C., Saurin W., Schlex T.,
  RA Sigler P., Thebaud P., Whalen M., Wincker P., Levy M.,
  RA Weissenbach J., Boucher C.A.;
  RT Genome sequence of the plant pathogen Ralstonia solanacearum";
  RL Nature 415:497-502(2002).
  CC -1- FUNCTION: Catalyzes the acyljoin condensation reaction between C
  atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
  1-deoxy-D-xylose-5-phosphate (DXP) (By similarity).
  CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-
  deoxy-D-xylose 5-phosphate + CO(2).
  CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By
  similarity).
  CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway, first step.
  CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
  step.
  CC -1- SUBUNIT: Homodimer (By similarity).

```

```

CC -1- SIMILARITY: Belongs to the transketolase family. DXS subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL: AL646068; CA015928.1; -.
DR HAMAP, MF_00315; -.
DR InterPro; IPR001017; Dehydrogenase_E1.
DR InterPro; IPR005477; Dxs.
DR InterPro; IPR005476; Transketolase_C.
DR InterPro; IPR005475; Transketolase_CR.
DR InterPro; IPR005474; Transketolase_N.
DR Pfam; PF00676; E1_dehydrog; 1.
DR Pfam; PF02779; transket_pyr; 1.
DR Pfam; PF02780; transketolase_C; 1.
DR TIGRFAMs; TIGR00204; dxs; 1.
DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
DR Transferrase; Flavoprotein; Thiamine pyrophosphate;
  KW Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
SQ
SEQUENCE 636 AA; 68300 MW; 819D1F42B9F1440B CMC64;

Query Match 62.5%; Score 2008.5; DB 1; Length 636;
Best Local Similarity 61.8%; Pred. No. 7.9e-134;
Matches 384; Conservative 92; Mismatches 134; Indels 11; Gaps 5;

QY 7 YELKNIHTPADIALSKDQLADEVRGYLTIVTSISGHPAAGLTVELTVAHYVF 66
DB 3 YELATTIDAPAEILRLDRQLGLADELRATVLSVAGTQGHLSNLTVELTVAHYVF 62
QY 67 NTPVDLVWDVGHQAYPEKILITGRKERPTIRTLGGVSAPFARDESDYAFGVGHSSTSI 126
DB 63 NTPDRIWVDVGHQGYPEKILITGRREGVATLRQLDGISGFRRSRSPDTEGTAAHSSTSI 122
QY 127 SAALGMALASQLSGDKKMAIIGDGSITGMAVEANAHAGDV-NANLTVILINDMSIS 185
DB 123 SAALGMALGACTQGNRAVAIVIGDMSAGMAFEANNAAGYRRLPLVYVILINDMSIS 182
QY 186 PYGANNVLTLYLSKFTSSVAREESKALAKRPSWELAKTHEEVKMINVDTLFEEL 245
DB 183 PYGALNNVLTLYLSKFTSSVAREESKALAKRPSWELAKTHEEVKMINVDTLFEEL 242
QY 246 GNYTGPIDGADVEMVSTLENKDL--LTG--PVGLHVTKKKGKGYAPAKEDPLAYHGP 301
DB 243 GNYTGPIDGADVEMVSTLENKDLNIRRALEGERPOLHVTKKKGKGYAPAKEDPLAYHGP 302
QY 302 APDPTKDFLPYAPSPHPTTEVFGRMLCDMAADERLLGITTPANREGSGIVFESQCFEN 361
DB 303 KENPDEGKIPARAPR-KVSYTVYRGOMLCDMAADKRLVGLTPARREGSGWVEFSRFTPD 361
QY 362 RYFDVALAEGHATVLAAGCQAGKAPVAIYSTLOSQYOLIHVALONLMDLFAIDRA 421
DB 362 RYFDVALAEGHATVLAAGCQAGKAPVAIYSTLOSQYOLIHVALONLMDLFAIDRA 421
QY 422 GLVPGDPTTHAGAFDYSYMRCPINMLIMAPADENECRQMLTTFQHHGPASVYRPGKGP 481
DB 422 GLVPGDPTTHAGAFDYSYMRCPINMLIMAPADENECRQMLTTFQHHGPASVYRPGKGP 481
QY 482 GAIDPTTTLAETGAETVRH-----HGRRIALLMGSWTFAVEAGKQAGTVVMRPFV 536
DB 482 GAIDPTTTLAETGAETVRH-----HGRRIALLMGSWTFAVEAGKQAGTVVMRPFV 536
QY 537 PPDQALVLEIARTHVPFTVEENVTAGAGSAINFTLOAKVLMPCNIGLDPREVEGGS 596
DB 542 PIDAAVLEIARTHVPFTVEENVTAGAGSAINFTLOAKVLMPCNIGLDPREVEGGS 601
QY 597 REELSLVGLDSKGLATIEQ 617

```

DB 602 HPAALACGLDANGILASIRE 622

RESULT 9

ID DXS_ECOLI STANDARD; PRT; 619 AA.

AC 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE 1-deoxy-D-xyulose 5-phosphate synthase (EC 2.2.1.7) (1-deoxyxyulose-5-phosphate synthase) (DXP synthase) (DXPS).

DN DXS OR B0420.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RA Bacterial.

RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RC STRAIN=K12;

RA MEDLINE=98151473; PubMed=9482846;

RA Lohs L.M., Campos N., Rosa Putra S., Danielsen K., Rohmer M., Boronat A.,

RT "Cloning and characterization of a gene from Escherichia coli encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxyulose 5-phosphate, a common precursor for isoprenoid, thiamin, and pyridoxol biosynthesis."

RT Proc. Natl. Acad. Sci. U.S.A. 95:2105-2110(1998).

RT [2]

RA SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RA MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis K.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.,

RT "The complete genome sequence of Escherichia coli K-12."

RT Science 277:1453-1474(1997).

RT [3]

RA SEQUENCE FROM N.A.

RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K., Duncan M., Federfpiel N., Hyman R., Kallman S., Komp C., Kurdi O., Lew H., Lin D., Nemeth A., Oefner P., Schramm P., Davis R.W., Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.

RT [4]

RA SEQUENCE OF 1-6, AND CHARACTERIZATION.

RA MEDLINE=98058734; PubMed=9371765;

RA Sprenger G.A., Schorken U., Wiesert T., Groll S., de Graaf A.A., Taylor S.V., Begley T.P., Bringer-Meyer S., Sahm H.,

RT "Identification of a thiamin-dependent synthase in Escherichia coli required for the formation of the 1-deoxy-D-xyulose 5-phosphate precursor for isoprenoids, thiamin, and pyridoxol."

RT Proc. Natl. Acad. Sci. U.S.A. 94:12857-12862(1997).

RT [5]

RA CHARACTERIZATION.

RA MEDLINE=20115529; PubMed=10648511;

RA Kuzuyama T., Takagi M., Takahashi S., Seto H.,

RT "Cloning and characterization of 1-deoxy-D-xyulose 5-phosphate synthase from Streptomyces sp. strain C190, which uses both the mevalonate and nonmevalonate pathways for isopentenyl diphosphate biosynthesis."

RT J. Bacteriol. 182:891-897(2000).

RT [6]

RA FUNCTION: Catalyzes the acylotin condensation reaction between C atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield 1-deoxy-D-xyulose-5-phosphate (DXP).

CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-deoxy-D-xyulose 5-phosphate + CO(2).

CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By similarity).

CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.

CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first step.

CC -1- SUBUNIT: Homodimer.

CC -1- MISCELLANEOUS: Optimal temperature is 42-44 degrees Celsius and optimal pH is 7.5-8.0.

CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF035440; AAC46162.1; -

DR EMBL; AF000148; AAC73523.1; -

DR EMBL; U82664; AAB40176.1; -

DR PIR; D64771; D64771.

DR EcoGene; E031612; dxs.

DR HAVAP; MF_003125; -1.

DR InterPro; IPR005477; Dxs.

DR InterPro; IPR005476; Transketolase_C.

DR InterPro; IPR005475; Transketolase_CR.

DR InterPro; IPR005474; Transketolase_N.

DR Pfam; PF02779; transket_pyr; 1.

DR Pfam; PF02780; transketolase_C; 1.

DR TIGRfams; TIGR00204; dxs; 1.

DR PROSITE; PS00801; TRANSKETOLASE_1; 1.

DR PROSITE; PS00802; TRANSKETOLASE_2; 1.

DR Transferrase; Flavoprotein; Thiamine pyrophosphate;

KW Isoprene biosynthesis; Thiamine biosynthesis; Manganese; Magnesium; Complete proteome.

FT INT MET

FT SEQUENCE 619 AA; 67465 MW; D1808BEEF9DC9D2B CRC64;

Query Match 60.2%; Score 2000.5; DB 1; Length 619;

Best Local Similarity 62.2%; Pred. No. 2-be-133;

Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

QY 7 YELKNIHTPADIDALSCKDQLOQLADVRGYLETFTVISGSHFAGLGTVELTVAAHYVF 66

DB 7 YETLALVDSTOELRLKRESIPKICDELRLYLDVSRRSHFASGGLTVELTVAAHYV 66

QY 67 MTPVDQVMDVGHQAYPEKILITGRKEMPTIRLTGVSAFPAPRESEYDARGVCHSSTSI 126

DB 67 NTPPDQILWDVGHQAYPEKILITGRKEMPTIRLTGVSAFPAPRESEYDARGVCHSSTSI 126

QY 127 SAALGMALASQLRGKEDKKMVALIGDSITGMAVEAYANNAAGDVANLVLINDMMSISP 186

DB 127 SAGGIAPAAKESKKNRTVCVIGDALTAKAFANNAHACDIPDMVLINDMMSISP 186

QY 187 PVGAMNNVLTVALSSKTYSSVREESKALAMPSEWELARTTEHYKGMIVGTLFEELG 246

DB 187 NVGALNNHLAGLISGLKYSSIREGKRVFSGVPIKEILKTEHYIKGMVVPGLFEELG 246

QY 247 FNYVGPIDGHVEMVNLSTLTXLXDLTGPVFLAVVTKGKGYPAPAKPDLAHPYAPEDPT 306

DB 247 FNYVGPIDGHVEMVNLSTLTXLXDLTGPVFLAVVTKGKGYPAPAKPDLAHPYAPEDPT 306

QY 307 KDFPKAPSPHPYTVTFVFGMLCDMAQDERLLGITPAMESSGLVFEPSKFPNRYTDV 366

DB 307 SGCLPKSS-GGLPFSYKIFGDMLCETAKDKMLAIFPAMESSGVMFEFSKFPDRYEDV 366

QY 367 AIAEHNATLAAGAQAQAGKRVNAIYSTFLQRGDOLIHVALNLDLFLPLDRAGLVGP 426

DB 367 AIAEHNATLAAGAQAQAGKRVNAIYSTFLQRGDOLIHVALNLDLFLPLDRAGLVGP 426

QY 427 DGPTTAGAFDSYMKCIENMLINAPADENECROMLTGFGQH-GRASVYRPGKPGGAI 485

DB 427 DGPTTAGAFDSYMKCIENMLINAPADENECROMLTGFGQH-GRASVYRPGKPGGAI 485

QY 426 DGGTHQAGFDSLAKCIPEAVITMPSDENECROMLYTGHYNDGSAVRYRGNAGVGL 485

DB 426 DGGTHQAGFDSLAKCIPEAVITMPSDENECROMLYTGHYNDGSAVRYRGNAGVGL 485

QY 486 DPTTLALBIGAEPVPHBSRLAILAMGSMVTPYAEAGSQLCATVNNRFPVDFDQALVLE 545

DB 486 DPTTLALBIGAEPVPHBSRLAILAMGSMVTPYAEAGSQLCATVNNRFPVDFDQALVLE 545

QY 486 TP-LEKLPIDGIVRREKRIALINFTLMPBAKVASINAVLDMKFAVPLDEALILE 544

DB 486 TP-LEKLPIDGIVRREKRIALINFTLMPBAKVASINAVLDMKFAVPLDEALILE 544

QY 546 LARTDVTVEENVYIAGAGSAINTPLOAKVMPVNCIGLPRFVEQSGREELISLVG 605
 DB 545 MAASHEALVTVEENVYIAGAGSAINTPLOAKVMPVNCIGLPRFVEQSGREELISLVG 604
 QY 606 LDSKGIATIBOCCA 620
 DB 605 LDAGMEAKIKAMLA 619

RESULT 10
 DKS_ECOL6 STANDARD; PRT; 619 AA.

AC 08FKB9;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate synthase (EC 2.2.1.7) (1-
 deoxyxyulose-5-phosphate synthase) (DXP synthase) (DXPS).
 DKS OR C0531.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=217992;
 [1]
 SEQUENCE FROM N.A.
 RP STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 Rasick D., Buckles E.L., Liu S.-R., Boutin A., Hackett J., Stroud D.,
 Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 CC -1- FUNCTION: Catalyzes the acyloln condensation reaction between C
 atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
 1-deoxy-D-xyulose-5-phosphate (DXP) (By similarity).
 CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-
 deoxy-D-xyulose 5-phosphate + CO(2).
 CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By
 similarity).
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.
 CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
 step.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EXBL: AEO16756; AAN79009.1; -.
 DR HAMAB: MF 00315; -; 1.
 DR InterPro: IPR005477; Dxs.
 DR InterPro: IPR005476; Transketolase_C.
 DR InterPro: IPR005475; Transketolase_CR.
 DR InterPro: IPR005474; Transketolase_N.
 DR Pfam: PF02780; transketolase_C; 1.
 DR Pfam: PF02779; transket_Pyr; 1.
 DR TIGRfam: TIGR00204; dks1; 1.
 DR PROSITE: PS00801; TRANSKETOLASE_1; 1.
 DR PROSITE: PS00802; TRANSKETOLASE_2; 1.
 RA Transferrase; Flavoprotein; Thiamine pyrophosphate;
 RA isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
 KW INIT MET 0 BY SIMILARITY.
 FT INIT MET 0 BY SIMILARITY.
 SQ SEQUENCE 619 AA; 67501 MW; E2A4AD1A564FC30 CRC64;

Query Match 62.2%; Score 1999.5; DB 1; Length 619;
 Best Local Similarity 60.0%; Pred. No. 3.3e-133;

Matches 369; Conservative 115; Mismatches 128; Indels 3; Gaps 3;
 QY 7 YPLKNIHTPADIRALSKDOQLQADENVGTLTHTVSISGHPAAGLGVETLVAHYV 66
 DB 7 YPTLALVDSQEHLLPKESLPKCDLRRYILDVSSRSSCHPASGIVETLVAHYV 66
 QY 67 NTPVDOLVMDVGHQAYHKKILITGKEMPTIRLTGSVAPAPADESRDAFGVSHSTSI 126
 DB 67 NTPVDOLVMDVGHQAYHKKILITGKEMPTIRLTGSVAPAPADESRDAFGVSHSTSI 126
 QY 127 SAALGMAIASQLRSDKQVALLIGDSITGMAEYANNAHDVANNILVILNDNMSISP 186
 DB 127 SAGIGIAVAAREKRRRTVCVIGDAITAGMAEYANNAHDVANNILVILNDNMSISP 186
 QY 187 PVGANNVLTIVISSKTVSYVRESKATALAKSVWELAKTEEHVKMIVPTLFEELG 246
 DB 187 NVGALNNHLLAQLLSGKLYSSLRBGKTVFSGVPIKEILKTEEHIKGMVPTLFEELG 246
 QY 247 FNYGPIDGHVEMVSTLENTKDLTGVPFLHVVTKKGVAPAPKDPAYHGVAPDPPT 306
 DB 247 FNYIGPVDSGHVGLITLTKNNRDLKGPFLHMTKRGTEPAEKDPITHAHPKDPDS 306
 QY 307 KDFLEKAPSPHPTTYTVFGRMLCMAAODERLIGITPAMEGSGLVESQKTPNRYDV 366
 DB 307 SGCLPKPS -GGLPSYSKLPEDMLCETAKDKMAITPAMEGSGMVEFSKFPDRYDV 365
 QY 367 ALAAGHAVTLAAGACGAKPVVAIYSTLTORGVDQLIHVVALQNTDMLPALDRAGLVGP 426
 DB 366 ALAAGHAVTLAAGALGKPVVAIYSTLTORGVDQLIHVVALQNTDMLPALDRAGLVGP 425
 QY 427 DGPTAGAFDYSYVRCIPNMLIMAPDENECROMLITGPGHH-GRASVYPRGKPGAAI 485
 DB 426 DGTHGAGFDLSYARCIPEVIMTPSDNECRQVLYGHNDPSAVRYRGNVGVEL 485
 QY 486 DPTTALRIGRAVYRHGSRITAIAMGSAVTPVAYEAGKQLGATVYNNRFPVPOATLE 545
 DB 486 TP-LEKPIGKIVYRRERKAIINFTLMEPAKVAESINATLVDMKVFVPLDLS 544
 QY 546 LARTDVTVEENVYIAGAGSAINTPLOAKVMPVNCIGLPRFVEQSGREELISLVG 605
 DB 545 MAASHEALVTVEENVYIAGAGSAINTPLOAKVMPVNCIGLPRFVEQSGREELISLVG 604

RESULT 11
 DKS_ECOL57 STANDARD; PRT; 619 AA.
 ID DKS_ECOL57
 AC 08XE76;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate synthase (EC 2.2.1.7) (1-
 deoxyxyulose-5-phosphate synthase) (DXP synthase) (DXPS).
 GN DKS OR Z0523 OR EC50474.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=83334;
 [1]
 SEQUENCE FROM N.A.
 RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 Grodzick E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,
 Apodaca J., Aantaahtaman T.S., Lin J., Yen G., Schwartz D.C.,
 Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RA Nature 409:529-533(2001).
 RL

RN SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / RIMD 0509952;
 RX MEDLINE=21156231, PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 Rida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RA "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1- FUNCTION: Catalyzes the acylol condensation reaction between C
 CC atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
 CC 1-deoxy-D-xylulose-5-phosphate (DXP) (By similarity).
 CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-
 CC deoxy-D-xylulose 5-phosphate + CO(2).
 CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By
 CC similarity).
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.
 CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
 CC step.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE005221; AAG54770.1; -;
 CC EMBL; AP002551; BAB33897.1; -;
 CC PIR; B90688; B90688.
 CC PIR; F85538; F85538.
 CC HAMAP: MF_00315; -; 1.
 CC InterPro: IPR005477; Dxs.
 CC InterPro: IPR005476; Transketolase_C.
 CC InterPro: IPR005475; Transketolase_CR.
 CC InterPro: IPR005474; Transketolase_N.
 CC Pfam: PF02779; transket pyr; 1.
 CC Pfam: PF02780; transketolase_C; 1.
 CC TIGRfam: TIGR00204; dxs; 1.
 CC DR TIGRfam: TIGR00204; dxs; 1.
 CC DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
 CC DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
 CC DR TRANSFERASE; Flavoprotein; Thiamine pyrophosphate;
 CC KW Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
 CC FT INIT MET 0 BY SIMILARITY.
 CC SQ SEQUENCE 619 AA; 67501 MW; F8150610E6FFC20 CRC64;
 Query Match 62.0%; Score 1995.5; DB 1; Length 619;
 Best Local Similarity 59.8%; Pred. No. 638-133;
 Matches 368; Conservative 115; Mismatches 129; Indels 3; Gaps 3;

Db 247 FNTGPGVDGHDVLTGTLITLTKNNRDLKGPQFLHMTKXGGRYPAEKDPITTHAVPKFDP 306
 Qy 307 KDFLRAPSPHPPTTEVFRMTCMDAAQBERLIGITPAKBSGGLVSEFQKPNRYFDV 366
 Db 307 SGLCPKSS -GGLPSTSTFEDMCTNADNKLMTTPNREBSGVNFRKRPDRFDV 365
 Qy 367 AIAEONAVTLAAGACGAGPVVAISTPLQRYDGIHDVALQNTDMLPALRAGIIVCP 426
 Db 366 AIAEONAVTLAAGIAGVPIVAISTPLQRYDGIHDVALQNTDMLPALRAGIIVCP 425
 Qy 427 DGPHTAGFPTSYMRCIPNMLIAPADENECRMLTTPGQSH -GPASVRYPRGKPAAI 485
 Db 426 DGPHTAGFPTSYMRCIPNMLIAPADENECRMLTTPGQSH -GPASVRYPRGKPAAI 485
 Qy 486 DPTVTALEIKARVHRHRSRIATLWGSVTPPAVEAGKQIGATVNNRFPKPDQALVE 545
 Db 486 TP-LEKPIKGIYKRGKELIIFGTLMPEAKVAESINATLVDRFPKPDLEITILE 544
 Qy 546 LARHDVPTVEERVIVGAGSAINTELOAKYMLPVCNIGLDRFVQSGREBELSLVG 605
 Db 545 MAASHVALTVEERVIVGAGSAINTELOAKYMLPVCNIGLDRFVQSGREBELSLVG 604
 Qy 606 LDSKILATIEOPCA 620
 Db 605 LDMAGMEAKIKAWIA 619
 RESULT 12
 Dxs_XANCP STANDARD; PRT; 638 AA.
 AC 08P815;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 1-deoxy-D-xylulose 5-phosphate synthase (BC 2.2.1.7) (1-
 DE deoxyxylulose-5-phosphate synthase) (Dxs synthase) (DXPS).
 GN Dxs OR XCC2434.
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 ON NCBI_TaxID=340;
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCPPB 528;
 RX MEDLINE=22022145; PubMed=12024217;
 RA de Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
 Camarotte G., Candavan F., Cardoso J., Chamego F., Ciapina L.P.,
 Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 Forughieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 Katayama A.M., Kisli L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 Locelli E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 Martins E.C., Meidanis U., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 Pereira H.A., Rossi A., Sema J.A.D., Silva C., de Souza R.F.,
 Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
 Setubal J.C., Kitajima J.P.;
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities.";
 RL Nature 417:459-463(2002).
 CC -1- FUNCTION: Catalyzes the acylol condensation reaction between C
 CC atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
 CC 1-deoxy-D-xylulose-5-phosphate (DXP) (By similarity).
 CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-
 CC deoxy-D-xylulose 5-phosphate + CO(2).
 CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By
 CC similarity).
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.
 CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
 CC step.

```

CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE012354; AAM41711.1; ALT_INIT.
CC HAMAP: MF_00315; -. 1.
CC InterPro: IPR005477; Dxs.
CC InterPro: IPR005476; Transketolase_C.
CC InterPro: IPR005475; Transketolase_CR.
CC InterPro: IPR005474; Transketolase_N.
CC Pfam: PF02779; transket_pyr_1.
CC Pfam: PF02780; transketolase_C_1.
CC TRFAMS: TRF00204; dxs; 1.
CC PROSITE: PS00801; TRANSKETOLASE_1; 1.
CC PROSITE: PS00802; TRANSKETOLASE_2; FALSE NEG.
CC Transferrase; Flavoprotein; Thiamine pyrophosphate;
CC Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
CC Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
CC SEQUENCE 638 AA; 68395 MW; AD53238C52FF739 CRC64;

Query Match          61.9%; Score 1989.5; DB 1; Length 638;
Best Local Similarity 61.4%; Pred. No. 1.7e-132;
Matches 378; Conservative 90; Mismatches 143; Indels 5; Gaps 3;

4 TTDYPLKNIHPADIRALSKQQLADBRGVLTHVTSISGHPAAGLTVLVAH 63
4 STYPRILSRIGTFDDIRFEEDLTVAVDELRAVLLIESVCKSGHFAAGLITLVLAH 63
4 YVNTPEVDLVVWVGVQAVYHKLITGRKEMPIRILGVSAPADSEGYAFGVGHS 123
64 YLQTPVDLVWVGVQAVYHKLITGRKEMPIRILGVSAPADSEGYAFGVGHS 123
124 TSISALGMAIALSOLGEGDKKVAIIIGDSITGMAVEAMNHG--DYANILVTLIND 181
124 TSISALGMAIALSOLGEGDKKVAIIIGDSITGMAVEAMNHG--DYANILVTLIND 183
124 TSISALGMAIALSOLGEGDKKVAIIIGDSITGMAVEAMNHG--DYANILVTLIND 183
182 MSISPPVGMANVITVLSKFSVSRRESKLA--KMSVVELARKTEEHKGVIVG 239
184 MSISEAVVGLITKMLGKSGSRITNALREGSKILGDKNNPVARFYRKEEHKGVIVG 243
240 TLFEEAGFNFGPIDGDMVEMVSTLENTKDLTGPFVFLVVTKKKGKGYAPAEKDPVLAHG 299
244 TLFEEAGFNFGPIDGDMVEMVSTLENTKDLTGPFVFLVVTKKKGKGYAPAEKDPVLAHG 303
300 VPAFDPDKDFLPRAADSPHPTTYEVFGRWLCMAAODERLLGITTPAKRSGSGIVRSQKF 359
304 VGFPEDSKGLVAKAG-AKKEPTVDVSDWVCKMAAADPKMLVITTPARBSGIVRSKEY 362
360 PNYVPEVAIEGHAVALTAAGACGAKPVAIYSTFGREYDULIDVALONTLMFLALD 419
363 PNYVPEVAIEGHAVALTAAGACGAKPVAIYSTFGREYDULIDVALONTLMFLALD 422
420 RAGLVGDPGPTHAADPYSYRCIPNNLINAADENECROMLTGQHHGPASVRYPRK 479
423 RGVVVGDPGATVHGNLDLSFLRCVPMVVA PADAEACROMLTGQHHGPASVRYPRK 482
480 GPGALDPTTALTEIGKAEVHGHSRIALAMGSMYTPAEAKKOLGATVNNRFKPP 539
483 GPGALDPTTALTEIGKAEVHGHSRIALAMGSMYTPAEAKKOLGATVNNRFKPP 542
540 QATVLELARTHDVFTVEENVVAGAGSAINTELQAKVLAPVCNIGLDPDFVEQSRRE 599
543 KAMLELAKGHEAVSIEDNVVAGAGSVELNAGVLAPEMLHGLPDSFGHARSRED 602
600 LLSIVELDSKGIATLTI 615
603 LLAENGIDQAGIRAAV 618

```

```

RESULT 13
DMS_XANAC STANDARD; PRT; 638 AA.
ID DMS_XANAC
AC Q8PUG7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 1-deoxy-D-xylulose 5-phosphate synthase (EC 2.2.1.7) (1-
DE 1-deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS).
CM DMS OR XAC2565.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Parah C.S., Furlan L.R.,
RA Quaglio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves I.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camarvan F., Cardoso J., Chamberg F., Ciapina L.P.,
RA Cicirelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Sepinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Setubal J.C., Kitchajima J.P.;
RA Trindade dos Santos M., Triffi D., Tsai S.M., White F.F.,
RT Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.
RL Nature 417:459-463(2002).
CC -1- FUNCTION: Catalyzes the acyloln condensation reaction between C
CC atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
CC 1-deoxy-D-xylulose-5-phosphate (DXP) (By similarity).
CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-
CC 1-deoxy-D-xylulose 5-phosphate + CO(2).
CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By
CC similarity).
CC -1- PATHWAY: Nucleotide biosynthesis; Thiamine biosynthesis; first step.
CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
CC step.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE011897; AAM37415.1; -.
CC HAMAP: MF_00315; -. 1.
CC InterPro: IPR005477; Dxs.
CC InterPro: IPR005476; Transketolase_C.
CC InterPro: IPR005475; Transketolase_CR.
CC InterPro: IPR005474; Transketolase_N.
CC Pfam: PF02779; transket_pyr_1.
CC Pfam: PF02780; transketolase_C_1.
CC TRFAMS: TRF00204; dxs; 1.
CC PROSITE: PS00801; TRANSKETOLASE_1; 1.
CC PROSITE: PS00802; TRANSKETOLASE_2; FALSE NEG.
CC Transferrase; Flavoprotein; Thiamine pyrophosphate;
CC Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
CC Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
CC SEQUENCE 638 AA; 68479 MW; 3D739EB912C17CF7 CRC64;

```

Query Match 61.8%; Score 1966.5; DB 1; Length 638;
 Best Local Similarity 61.4%; Pred. No. 2.8e-132;
 Matches 378; Conservative 91; Mismatches 142; Indels 5; Gaps 3;

4 TTDYPLKNIHTPADIRALSKDQLOQLADREYGLTHVTISISGHPAAGVGLVTLVLAH 63
 4 TTYPRLSRIQTDPDDRPEAEITLAEELRSYLIESVSGGHPAAGVGLVTLVLAH 63
 4 TTYPRLSRIQTDPDDRPEAEITLAEELRSYLIESVSGGHPAAGVGLVTLVLAH 63
 64 YVNTPTDQVWVNGVQVAPKHLITGKEMPTIRLGGVSAIPADDESIVDFGSHS 123
 64 YVNTPTDQVWVNGVQVAPKHLITGKEMPTIRLGGVSAIPADDESIVDFGSHS 123
 64 YVNTPTDQVWVNGVQVAPKHLITGKEMPTIRLGGVSAIPADDESIVDFGSHS 123
 124 TSISALGMAIASOLREDEKKNVAILGDSITGMAVEANMHG--DVANLTVIINDMD 181
 124 TSISALGMAIASOLREDEKKNVAILGDSITGMAVEANMHG--DVANLTVIINDMD 181
 124 TSISALGMAIASOLREDEKKNVAILGDSITGMAVEANMHG--DVANLTVIINDMD 181
 182 MSISPPVGMANNVITKYLSSKFTSVSEESKALA--KMSVWELAKTEEVYKAVIPG 239
 184 MSISAVGGLTKMLGRASGSRITLAIKGGKILGDKNNPTARFYRWEHEWKGAFVPS 243
 240 TLEHIGVNYFGPIGDGVEMVLTENKDLGVPVFLVYVTKKGYAPAEADPLAHS 299
 244 TLEHIGVNYFGPIGDGVEMVLTENKDLGVPVFLVYVTKKGYAPAEADPLAHS 299
 300 VPAPDPKDELPRKAPSPHPTTEVFGRMLCMAAODERLLGITPARRSGSGVSESOKE 359
 304 VGPDPKDELPRKAPSPHPTTEVFGRMLCMAAODERLLGITPARRSGSGVSESOKE 362
 360 PNTYFVAIAEQAHTVLAAGACGAGVVAISTTQGGYDGLHVALQNTDMLPADD 419
 363 PNTYFVAIAEQAHTVLAAGACGAGVVAISTTQGGYDGLHVALQNTDMLPADD 422
 420 RAGLVGDPGPHAPAFYSYVRCIPNMLIAPADENRCROMLTGSGHHPAARVPRGK 479
 423 RAGLVGDPGPHAPAFYSYVRCIPNMLIAPADENRCROMLTGSGHHPAARVPRGK 482
 483 RAGLVGDPGPHAPAFYSYVRCIPNMLIAPADENRCROMLTGSGHHPAARVPRGK 482
 480 RAGLVGDPGPHAPAFYSYVRCIPNMLIAPADENRCROMLTGSGHHPAARVPRGK 479
 483 RAGLVGDPGPHAPAFYSYVRCIPNMLIAPADENRCROMLTGSGHHPAARVPRGK 482
 540 OALVLEHARTDVTVEENVVAGAGSALINTPLQAKVIMPVONISLPPRFVSGSREE 599
 543 KALLELEAKTHDGVTTLEDVNVAGAGSVALINAPALITPLHLGLDPSFQHHASRED 602
 600 LSLVGLDSKGLIATI 615
 603 LLAERAGIDQAGIRAL 618
 RESULT 14
 DLS_PASMTU STANDARD; PRT; 614 AA.
 AC P57848;
 DT 16-OCT-2001 (Rel. 40, Last Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.1.7) (1-
 deoxyxylose-5-phosphate synthase) (DXP synthase) (DXPS).
 GN DXP OR PM0532.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida Pm70."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 CC -1- FUNCTION: Catalyzes the acyloin condensation reaction between C
 atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield

CC 1-deoxy-D-xylose-5-phosphate (DXP) (By similarity).
 CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-
 CC deoxy-D-xylose 5-phosphate + CO(2).
 CC -1- CORFACTOR: Binds 1 thiamine pyrophosphate per subunit (By
 CC similarity).
 CC -1- PATHWAY: Nucleoside triphosphate biosynthesis pathway; first step.
 CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
 CC step.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AE006088; AAK02616.1; -.
 DR HAMAP; MF 00315; -; 1.
 DR InterPro; IPR005477; Dxa.
 DR InterPro; IPR005476; Transketolase_C.
 DR InterPro; IPR005475; Transketolase_CR.
 DR InterPro; IPR005474; Transketolase_N.
 DR Pfam; PF02779; transket_pyr; 1.
 DR Pfam; PF02780; transketolase_C; 1.
 DR TIGRfam; TIGR00204; dxa; 1.
 DR PROSITE; PS00601; TRANSKETOLASE_1; 1.
 DR PROSITE; PS00602; TRANSKETOLASE_2; 1.
 DR Transferrase; Flavoprotein; Thiamine pyrophosphate; Complete
 KM isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
 SQ SEQUENCE 614 AA; 67426 MW; E85D95DBBCAD3 CR64;
 Query Match 61.2%; Score 1969; DB 1; Length 614;
 Best Local Similarity 60.1%; Pred. No. 4.5e-131;
 Matches 369; Conservative 108; Mismatches 133; Indels 4; Gaps 3;

6 DYPLKNIHTPADIRALSKDQLOQLADREYGLTHVTISISGHPAAGVGLVTLVLAH 65
 3 NYPLSLINSPELDLILSKEDLPQICORLAVLSSVSSGHLASGLATVLAH 62
 66 FNTPVQOLVWDVGHQVAPKHLITGKEMPTIRLGGVSAIPADDESIVDFGSHS 125
 63 YVNTPTDQVWVNGVQVAPKHLITGKEMPTIRLGGVSAIPADDESIVDFGSHS 122
 126 ISALGMAIASOLREDEKKNVAILGDSITGMAVEANMHG--DVANLTVIINDMD 185
 123 ISALGMAIASOLREDEKKNVAILGDSITGMAVEANMHG--DVANLTVIINDMD 182
 186 PNTYFVAIAEQAHTVLAAGACGAGVVAISTTQGGYDGLHVALQNTDMLPADD 419
 183 PNTYFVAIAEQAHTVLAAGACGAGVVAISTTQGGYDGLHVALQNTDMLPADD 422
 245 OALVLEHARTDVTVEENVVAGAGSALINTPLQAKVIMPVONISLPPRFVSGSREE 599
 243 KALLELEAKTHDGVTTLEDVNVAGAGSVALINAPALITPLHLGLDPSFQHHASRED 602
 305 LSLVGLDSKGLIATI 615
 303 LLAERAGIDQAGIRAL 618
 365 DVAIAEQAHTVLAAGACGAGVVAISTTQGGYDGLHVALQNTDMLPADD 419
 361 DVAIAEQAHTVLAAGACGAGVVAISTTQGGYDGLHVALQNTDMLPADD 422
 425 GPDGPHAPAFYSYVRCIPNMLIAPADENRCROMLTGSGHHPAARVPRGK 479
 421 GPDGPHAPAFYSYVRCIPNMLIAPADENRCROMLTGSGHHPAARVPRGK 482
 485 IDTTLALGIAEQAHTVLAAGACGAGVVAISTTQGGYDGLHVALQNTDMLPADD 419
 481 IDTTLALGIAEQAHTVLAAGACGAGVVAISTTQGGYDGLHVALQNTDMLPADD 422

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:14 ; Search time 46.2142 Seconds
(without alignments)
3461.979 Million cell updates/sec

Title: US-09-941-947a-6

Sequence: 1 MKLTTDFPLKXNHTPADIR.....LSLVGLDSKGLATIEQCA 620

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

SPTRMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriag.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2114	65.7	621	16	Q8DF33 vibrio vuln
2	2042	63.5	622	16	Q8EGR9 shewanella
3	1999.5	62.2	620	16	Q8FKB9 escherichia
4	1795	55.8	626	16	Q8D357 wiggleswort
5	1592.5	49.5	643	16	Q8G292 brucella su
6	1498.5	46.6	644	16	Q8KFI9 chlorobium
7	1482.5	46.1	716	10	Q8D174 synechococc
8	1445	45.2	655	16	Q8D174 synechococc
9	1444.5	44.3	711	10	Q8I692 medicago tr
10	1440.5	44.8	715	10	Q8I692 medicago tr
11	1437.5	44.7	630	2	Q8K213 unclutered
12	1435.5	44.6	709	10	Q8FSG2 narctissus p
13	1435.5	44.6	713	10	Q8PD65 atemistia a
14	1433	44.6	736	2	Q8GAA0 synechococc
15	1430.5	44.5	732	2	Q8W4V1 morinda cit
16	1429.5	44.4	717	10	Q8I693 medicago tr

17	1418.5	44.1	724	10	Q64904 mentha pipe
18	1408.5	43.8	725	10	Q9FV39 tagetes ere
19	1402.5	43.6	719	10	Q9XHS0 lycopersico
20	1402	43.5	735	10	Q8I954 chlamydomon
21	1397.5	43.5	719	8	Q78328 capsicum an
22	1311.5	40.8	604	10	Q8I9F5 arabidopsis
23	1278.5	39.8	628	10	Q8L9S4 arabidopsis
24	1235.5	38.4	634	16	Q8FI53 leptospira
25	1131.5	35.2	616	16	Q8FPI2 coriynobact
26	1111.5	34.6	649	2	Q9FIV2 kitasatospo
27	1097	34.1	633	2	Q8VUR8 kitasatospo
28	1078.5	33.5	642	16	Q8CJP7 streptomyce
29	1062.5	33.0	628	10	Q9SK01 oryza sativ
30	1053	32.7	700	10	Q9LEF9 arabidopsis
31	1030	32.0	600	16	Q8R639 fusobacteri
32	892.5	27.8	586	16	Q97J35 clostridium
33	886	27.5	580	16	Q9CF08 lactococcus
34	882.5	27.4	599	16	Q9CRJ5 lactococcus
35	870	27.1	583	16	Q8R606 fusobacteri
36	854	26.6	354	2	Q89774 roseobacter
37	852.5	26.5	336	16	Q50408 mycobacteri
38	848.5	26.4	428	10	Q944G7 arabidopsis
39	823	25.6	406	10	Q94CE7 arabidopsis
40	819.5	25.5	1205	5	Q96694 plasmodium
41	819.5	25.5	1205	5	Q8IDM0 plasmodium
42	803.5	25.0	612	16	Q8EXK7 mycoplasma
43	689	21.4	774	10	Q8G585 bifidobacte
44	615	19.1	294	10	Q9FEV6 nicotiana t
45	522	16.2	272	2	Q9WXP0 dehnococcus

ALIGNMENTS

RESULT 1

ID Q8DF33 PRELIMINARY; PRT; 621 AA.
AC Q8DF33:
DT 01-MAR-2003 (TREMBLrel_23, Created)
DT 01-MAR-2003 (TREMBLrel_23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel_23, Last annotation update)
DE Deoxyxylulose-5-phosphate synthase.
GN VVI0315.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_taxid=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhoe J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.,
RT "Complete genome sequence of *Vibrio vulnificus* CMCP6";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016798; AAC08845.1; -
KW Complete proteome.
SQ SEQUENCE 621 AA; 67959 MM; 69DABA88B228D4B CMC64;

QY	3	LYTTD---YELKXNHTPADIRALSKDQLOQLADEVRGYLHTVTSISGGHAPAGLCTVELT	59
DB	1	MTLIDSKPPLALAEETPOHRLPLRETLPTLCDRLRYLANSVSQSSGHLASGLTVELT	60
QY	60	VALHVFVPTPDQLVMPYVGHQAYPHKLTGKEMPTIRLTGGVSAFAPARDESEYDAGV	119
DB	61	VALHVFVPTPDQLVMPYVGHQAYPHKLTGKEMPTIRLTGGVSAFAPARDESEYDAGV	120
QY	120	GHSSSTISALGMAISQIRGEDKRWVAIGDSITGMAVEAMNHADVNAALVTLND	179
DB	121	GHSSSTISALGMAISQIRGEDKRWVAIGDSITGMAVEAMNHADVNAALVTLND	180

QY 180 NMDSISPPVGMANNLTITVLSKFSYSSREESKALAKMPVWEIARTTEEHKMGTVRG 239
 DB 181 NMDSISPPVGMANNLTITVLSKFSYSSREESKALAKMPVWEIARTTEEHKMGTVRG 240
 QY 240 TLFEELGPNVIGPIGHDVEMVSTLENLKDTLGVFHVTVTKKGGYAPAEKDPYAVHG 299
 DB 241 TLFEELGPNVIGPIGHDVEMVSTLENLKDTLGVFHVTVTKKGGYAPAEKDPYAVHG 300
 QY 300 VPAFPTKDFLPKAPSPHPTTTEVFGKRLCDMAADDERLLGTPAMRESGGLVPSQKF 359
 DB 301 VPAFPTKDFLPKAPSPHPTTTEVFGKRLCDMAADDERLLGTPAMRESGGLVPSQKF 359
 QY 360 PRRYDVDAIAEQHATVLAAGACOGAKPVVAIYSTFLORGVDOLIHVALONLMLFALD 419
 DB 360 PRRYDVDAIAEQHATVLAAGACOGAKPVVAIYSTFLORGVDOLIHVALONLMLFALD 419
 QY 420 RAGLVGPDGPTHAGAPDVSVMCIENMLIMADENECRQMLTTFQHHGPASVYPRGK 479
 DB 420 RAGLVGPDGPTHAGAPDVSVMCIENMLIMADENECRQMLTTFQHHGPASVYPRGK 479
 QY 480 GGGTPIESFTLEIGKRLVQSEKVALSFTGLANLBAEALNATVADMRFVRLD 539
 DB 480 GGGTPIESFTLEIGKRLVQSEKVALSFTGLANLBAEALNATVADMRFVRLD 539
 QY 540 QALVLELATHDVFVTEENVLAGAGSAINTELQAKVLPVCNIGLPRFVQSGSEEE 599
 DB 540 QALVLELATHDVFVTEENVLAGAGSAINTELQAKVLPVCNIGLPRFVQSGSEEE 599
 QY 600 LSLVGLDSKGLATITRQCA 620
 DB 600 LSLVGLDSKGLATITRQCA 620

RESULT 2
 Q8EB9 PRELIMINARY; PRT; 622 AA.
 AC Q8EB9; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE deoxyxylulose-5-phosphate synthase.
 GN DKS OR S01525.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OC NCBI_Taxid=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NR-1.
 RX MEDLINE=2237686; PubMed=12368813;
 RA Heideberg J.F., Paulsen I.T., Nelson K.E., Galois E.J., Nelson W.C.,
 Read T.D., Bisen J.A., Seshadri R., Ward N., Mehe B., Clayton R.A.,
 Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Impriali M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouzari H., Gill J., Uutterback T.R., McDonald L.A.,
 RA Feldlyum T.V., Smith H.O., Venter J.C., Neillson K.H., Fraser C.M.,
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 Shewanella oneidensis."
 RL Nat. Biotechnol. 20:1118-1123 (2002).
 DR EMBL; AB015598; AAN54586.1; -.
 DR TIGR; S01525; -.
 KW Complete proteome.
 SQ SEQUENCE 622 AA; 68093 MM; BE681C01B8A5EB4 CRC64;

Query Match 63.5%; Score 2042; DB 16; Length 622;
 Best Local Similarity 62.0%; Pred. No. 2.5e-140;
 Matches 303; Conservative 105; Mismatches 125; Indels 4; Gaps 3;
 5 TDYPLAKHTPTDIALSKDQOQGLADVRGLTHTVTSISGHPAAGTVELTVALHY 64

DB 6 SDFVLAQANTPNELRQLPOLLADELREELINKSGMSGHFASGLQVELTVALHY 65
 QY 65 VFNTPVDQVWDVGHQAPPKKLTGKERMPRTITLGVSAPAPADSEYDAFGVST 124
 DB 66 VFNTPVDQVWDVGHQAPPKKLTGKERMPRTITLGVSAPAPADSEYDAFGVST 125
 QY 125 SISALMALASQURGEDEKMAV1IGD3ITGMAVAVANNAADVANNILVINDMS1 184
 DB 126 SISALMALASQURGEDEKMAV1IGD3ITGMAVAVANNAADVANNILVINDMS1 185
 QY 185 SPPVGMANNLTITVLSKFSYSSREESKALAKMPVWEIARTTEEHKMGTVRG 244
 DB 186 SPPVGMANNLTITVLSKFSYSSREESKALAKMPVWEIARTTEEHKMGTVRG 245
 QY 245 LGFVFGPIGHDVEMVSTLENLKDTLGVFHVTVTKKGGYAPAEKDPYAVHG 304
 DB 246 LGFVFGPIGHDVEMVSTLENLKDTLGVFHVTVTKKGGYAPAEKDPYAVHG 305
 QY 305 PTKDFLPKAPSPHPTTTEVFGKRLCDMAADDERLLGTPAMRESGGLVPSQKF 363
 DB 306 PTKDFLPKAPSPHPTTTEVFGKRLCDMAADDERLLGTPAMRESGGLVPSQKF 363
 QY 364 FDVALAEOHATVLAAGACOGAKPVVAIYSTFLORGVDOLIHVALONLMLFALD 423
 DB 364 FDVALAEOHATVLAAGACOGAKPVVAIYSTFLORGVDOLIHVALONLMLFALD 423
 QY 424 VGGDPTHAGAPDVSVMCIENMLIMADENECRQMLTTFQHHGPASVYPRGK 482
 DB 424 VGGDPTHAGAPDVSVMCIENMLIMADENECRQMLTTFQHHGPASVYPRGK 483
 QY 482 VGGDPTHAGAPDVSVMCIENMLIMADENECRQMLTTFQHHGPASVYPRGK 483
 DB 483 VGGDPTHAGAPDVSVMCIENMLIMADENECRQMLTTFQHHGPASVYPRGK 483
 QY 483 AALPPTLTALEIGABVRHSGRLAILAGSMVTPAVEAGKOLGATVNNRFPVPOAL 542
 DB 484 AALPPTLTALEIGABVRHSGRLAILAGSMVTPAVEAGKOLGATVNNRFPVPOAL 543
 QY 543 VLELATHDVFVTEENVLAGAGSAINTELQAKVLPVCNIGLPRFVQSGSEEE 602
 DB 544 VLELATHDVFVTEENVLAGAGSAINTELQAKVLPVCNIGLPRFVQSGSEEE 603
 QY 603 LVGLDSKGLATITRQCA 620
 DB 604 LVGLDSKGLATITRQCA 621

RESULT 3
 Q8EB9 PRELIMINARY; PRT; 620 AA.
 AC Q8EB9; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE 1-deoxy-D-xylulose 5-phosphate synthase (EC 4.1.3.37).
 GN DKS OR C0531.
 OS Bacterioides coli 06.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Bacterioides.
 OC NCBI_Taxid=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=06:HI / CPT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 DR EMBL; AB016756; AAN79009.1; -.
 KW Complete proteome.
 SQ SEQUENCE 620 AA; 67633 MM; D9DCB76CA65BA7BA CRC64;

Query Match 62.2%; Score 1999.5; DB 16; Length 620;
 Best Local Similarity 60.0%; Pred. No. 3.2e-137;

Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).
 DR EMBL: AE014353; AAN29379.1;--
 DR TIGR: BR0436;--
 KW Complete proteome.
 SQ SEQUENCE 643 AA; 69179 MW; 342DFB422D129FEA CRC64;

Query Match 49.5%; Score 1592.5; DB 16; length 643;
 Best Local Similarity 51.4%; Pred. No. 1.7e-107;
 Matches 318; Conservative 107; Mismatches 181; Indels 13; Gaps 7;

```

QY 8 PLKNIHTPADIRALSKOOLQOLADVEVGYLTHVTSISGHPAAGTVELTVALHYEN 67
DB 7 PLDKRAPPPDRALRPEODLPQLAEELTELIDAVSTTGSHLGLVVELTVALHYEN 66
QY 68 TPVDOLVMDVGHQAVPHKILITGSKERMPITRTLGVSAPFAPADESEYDAFGVSHSSTIS 127
DB 67 TPYDRIIMDVGHQAVPHKILITGSKERMPITRTLRIRITLQAGLSGFTGPAESEYDPFGAASHSTIS 126
QY 128 AALGMAIASQLRGEDEKKNVAITIGDSITGGAAVEANNAHADVANNALVTLINDMKSISP 187
DB 127 AGLGMAVAVASELSGKRNVAIVIGDSMSKGAAYEAMNNAAGLDAELIVTLINDMKSIAIP 186
QY 188 VGAMNNVYLTIVLSKPYSSVRESKKAALAKMPS-VWELARKTEBRYKGIIVPTLFEELG 246
DB 187 TGAASVYLAARLVSGRTYRSVRBAKQVAKLEKPTQDPAKRSSEYARAFPTGTLFEELG 246
QY 247 FNYFPGIDGHVEMLVSTLENLKD-TGPFVFLAVVTKKGGYAPAEKDPFLATRGVAPADP 305
DB 247 FYYVGGIDGHNDHLLPVTKNVRDTQKGPVLIHVVTQKGGYAPAAADAKHYGNKEDV 306
QY 306 TPDPLPKAPSPHPPTTYTVEGKMLCDMAODERLIGTPAMESSGSGIVEFSQKPEPNRFPD 365
DB 307 ITGKQAK-PPAVNPSTYKIFGTSLTBRAHDKIAVTAAMPETGGLDGEAFKRYFD 365
QY 366 VALAEQHAVTTLAAGACQAGKPVVAIYSTELQRYDQLIHDVALQNLDMLEPALDAGLVG 425
DB 366 VGIABGHATTPAAGLASSEYKPCALYSTELQRYDQVHVDSIQNLVPRFPIDRAGLVG 425
QY 426 PPGPTTAGAFDYSYKRCIPNMLIMAPADENEGQMLTTFQH-HGPASRYTRGKGPAA 484
DB 426 AGGPPTAGFDTGFLAALPGFVMAASDAEELHMYRTAEYDEGTSIFRRYRGDGVGD 485
QY 485 IDPTLTALITGKAEVHHGSRITAILAMGSMTPPAVEAGKOLG-----TVNMRFEKPED 539
DB 486 LBERGSLVITGKCRITRREGTKALLSFRRLDQCLAAAEILGAAGSTVPAARAKPLD 545
QY 540 QALVLELARTHDVFTVEBNVITAGAGSAINFLQAKVLA---MPCNIGLPRDFVQGS 596
DB 546 HDLIRLAREHEVLVWVEGAV-GFGSHVLQFLATDGLDRGLKVRALLTLPDIYQDHGK 604
QY 597 REELSLVGLDSKGLIATI 615
DB 605 PDAMVAEAGLDRTGIVRTV 623

```

RESULT 6
 O8KF19 PRELIMINARY; PRT; 644 AA.
 AC O8KF19, 01-OCT-2002 (TEMBLrel. 22, Created)
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE 1-deoxyxylulose-5-phosphate synthase.
 GN DKS OR CT0337.
 OS Chlorobium tepidum
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
 OC Chlorobium.
 NCBI Taxid=1097;
 OX (1)
 RN SEQUENCE FROM N.A.
 RP STRAIN=JLS / ATCC 49652 / DSM 12025;
 RX MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,

RA Dodson R.J., Deboy R., Gwim M.L., Nelson W.C., Haft D.H.,
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
 RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
 RA Nieman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Rutledge D.,
 RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M., Parksey D.,
 RT "The complete genome sequence of Chlorobium tepidum, a
 RT photosynthetic, anaerobic, green-sulfur bacterium."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).
 DR EMBL: AE012812; AAM71583.1;--
 DR TIGR: CT0337;--
 DR InterPro: IPR005477; Dxs.
 DR InterPro: IPR005476; Transketolase_C.
 DR InterPro: IPR005475; Transketolase_CR.
 DR InterPro: IPR005474; Transketolase_N.
 DR Pfam: PF02780; transketolase_C_1.
 DR Pfam: PF02779; transket. pyr. 1.
 DR TIGRFAMs: TIGR00204; dks; 1.
 DR PROSITE: PS00801; TRANSKETOLASE_1; 1.
 KW Complete proteome.
 SQ SEQUENCE 644 AA; 69739 MW; DE2FB8181AF2B63 CRC64;

Query Match 46.6%; Score 1498.5; DB 16; length 644;
 Best Local Similarity 49.3%; Pred. No. 1.3e-100;
 Matches 308; Conservative 93; Mismatches 211; Indels 13; Gaps 6;

```

QY 7 YPLKNIHTPADIRALSKOOLQOLADVEVGYLTHVTSISGHPAAGTVELTVALHYE 66
DB 15 YPLSLSHSPADLKCSIHLELVAACRKKVIELVSQNGHSSSGVVELTVALHYV 74
QY 67 NTPVDOLVMDVGHQAVPHKILITGSKERMPITRTLGVSAPFAPADESEYDAFGVSHSSTI 126
DB 75 QSPDRIIMDVGHQAVPHKILITGSLAMETNRHGLAGPKESEPHDAGTGHASTSI 134
QY 127 SAALGMAIASQLRGEDEKKNVAITIGDSITGGAAVEANNAHADVANNALVTLINDMKSISP 186
DB 135 SAAAGLAAARLALARKCKKVAIIGDSITGGAAVEANNAHADVANNALVTLINDMKSISP 194
QY 187 PVGAMNNVYLTIVLSKPYSSVRESKKAALAKMPS-VWELAR-----KTEBRYKGIIVPTL 241
DB 195 STGGKKNLVNLTAKTYNRLKRVWPSLSLHNEIGETAKTANHRIEDGKAFPTPAY 254
QY 242 FEELGFNYFPGIDGHVEMLVSTLENLKD-TGPFVFLAVVTKKGGYAPAEKDPFLATRGV 300
DB 245 FEALGFRYFGPIDGHNDHLLPVTKNVRDTQKGPVLIHVVTQKGGYAPAAADAKHYGNKEDV 314
QY 301 PAPPTDPLPKAPSP-HPPTTYTVEGKMLCDMAODERLIGTPAMESSGSGIVEFSQK 359
DB 315 GGFPIETGKNVKAQKPKPKTQEVFGALVELLKPPTITATLAAHPSGSLDLFQCAI 374
QY 360 PNRYFDVALAEQHAVTTLAAGACQAGKPVVAIYSTELQRYDQLIHDVALQNLDMLEPALD 419
DB 375 PSRCIDVGLAQHVAHTPAAGLASSEYKPCALYSTELQRYDQVHVDSIQNLVPRFPIDRAGLV 434
QY 420 PPGPTTAGAFDYSYKRCIPNMLIMAPADENEGQMLTTFQH-HGPASRYTRGKGPAA 478
DB 425 RAGLVGDEGPTTHGAFDLSYLVNVPNLITAPGDBQELRNLTALYDIPKPAIRPRG 494
QY 479 KQPGALIDPTLTALITGKAEVHHGSRITAILAMGSMTPPAVEAGKOLG-----TVNMR 533
DB 485 SCSGATLHKETPLPVGVGRILRDGKSVALLGITMSRBALETAALGAAGLDPVCMR 554
QY 534 TVKPPDQALVLELARTHDVFTVEBNVITAGAGSAINFLQAKVLAFCVCHIGLPRDVE 593
DB 555 FLKPLDTEITIDMAASRCHTIYTIENSITIGFSGNVVYIHAHFGIKCISFGLPDAVT 614
QY 594 QGSREELSLVGLDSKGLIATIBOP 618
DB 615 HGSMDLIRVGLDARSLSKGLIEF 639

```

RESULT 7
 O82676

ID 082676 PRELIMINARY; PRT; 716 AA.
 AC 082676;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DB 1-deoxy-xylose 5-phosphate synthase.
 OS Catharantus roseus (Rosy periwinkle).
 DE Catharantus roseus (Rosy periwinkle).
 CC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asteridae; Lamiales; Gentianales; Apocynaceae; Rauvolfiaceae; Vinnaceae;
 CC Catharantus.
 NCBI_TaxID=4058;
 RN NCBI_TaxID=4058;
 RP SEQUENCE FROM N.A.
 RA Chahed K., Oudin A., Guivarc'h N., Hamdi S., Chentleux J.C., Rideau M.,
 RA Clastre M.;
 RL "1-deoxy-D-xylose 5-phosphate synthase from periwinkle: cDNA
 RT identification and induced gene expression in terpenoidindole
 RT alkaloid-producing cells.";
 RL Plant Physiol. Biochem. 38:559-566 (2000).
 DR EMBL; AL011840; CA09804.2; -.
 DR InterPro; IPR005476; Transketolase_C.
 DR InterPro; IPR005475; Transketolase_CR.
 DR Pfam; PF02780; transketolase_C.1.
 DR TIGRFAMs; TIGR00204; dxs; 1.
 DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
 SQ SEQUENCE 716 AA; 76793 MW; E5D75ED703246F7D CRC64;

Query Match 46.1%; Score 1482.5; DB 10; Length 716;
 Best Local Similarity 48.3%; Pred. No. 2.2e-99;
 Matches 307; Conservative 109; Mismatches 188; Indels 31; Gaps 11;
 8 PLKNIHTPADIRALSKDLOQLADVEGYLTHTVYISGSHFAAGTVELTVALHYEYFN 67
 72 ILIDITINYVHKNTLSHDLLEQLAEIRAEIYSAKVGSHLSLGVVDLTVALHVFN 131
 68 TPVDOLWVDVGHQAYPHKILITGRKERMPITRTLGVSAPPADESEYDAFGVSHSTIS 127
 132 TPEDRIWVGHQAYPHKILITGRKERMPITRTLGVSAPPADESEYDAFGVSHSTIS 191
 128 ALGALIASQLRGEDEKQVAIIIGDSITGMAVEAMNNAHDVNAHLVTLINDMSISP 186
 192 AGLGAAVADIIIGKNNVSVIGDGAHTAGQAYEAMNNAHFLDNLIVLINDNKVSLPT 251
 187 -----PVGAMNNYLTIVLSKPYSSVRESKALAKM-PSWELARKTEEHVKMI- 236
 252 ATLDDGATFVGLSALSLKQSPKREKRAKSIITKOIGFAHEVAAKVDEYARGLS 311
 237 -VPGITFELGNNYRPGIDGHVEMLVSTLENLKL--TGPFVLTWTKKGGVAPAKD 293
 312 ATGSTLFEELGIYYIGPVDSHIEDLVITFQYKAPAPGVPVLIHIVKGGSPPARVA 371
 294 PLAYHGVPAFDPPT--KDFLPKAAPEPHPTVEFGRLCDMAADERLIGITPARESGS 351
 372 ADRMGVAVFDPKGTGQFSK---SPTLSITYFABSLTKBAIDMKIIAIHMAAGGSG 428
 352 LVEBSQKPNRYFDVALAQHAYTLAAGQACQAKVVAIYSTFLQRGDOLIHVALON 411
 429 LNYFQKRPDRCGDVGLAIOHAYTPAAGLATBGLKPCAIIYSSFLQRGDQVHVHDLQK 488
 412 LMDLFLADAGLVGPGPTHAGAFDYSYRCIPNNLIMAPADENECROWLTGFP--HGP 470
 483 LPRFAMDAGLVGADGFTHCAGFDVAVACLPNMIIVAPSDBAELMHVAVAPAKIDRP 548
 471 ASVRYRGKGPAAIDPTL--TALIGKAEVRHSGRIIILAMGSVTPPAVAKOL--- 525
 549 CCRFRPRGNGIGVALPNNKGTPLBEGKRIIVGSGRAVILIGYSTVQDCLAAEMLKSH 608
 526 --GATVNNRFTYKPEQCALVLELARTHDVYVTEENVVAGGASALNTEFGAKVLA--M 580
 609 NNSPTVADAKFCPLDGDILKTLAKHEHILITVEBSI--GGFSGVTHLSTLIGLIGRI 667

Query 581 PYCNIGLPRFVEQSGREBLISVGLDSKGIATL 615
 668 KVASLFLPRYIDHGAAPVDQIEAGLSRHICATI 702

RESULT 8

ID 08DL74 PRELIMINARY; PRT; 655 AA.
 AC 08DL74;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE 1-deoxy-xylose 5-phosphate synthase.
 GN TIL0623.
 OS Synechococcus elongatus (Thermosynechococcus elongatus).
 CC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 NCBI_TaxID=32046;
 RN NCBI_TaxID=32046;
 RP SEQUENCE FROM N.A.
 RA STEALIN-BP-1;
 RX MEDLINE=22225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shingo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynechococcus elongatus BP-1.";
 RL DNA Res. 9:123-130 (2002).
 DR EMBL; AP005371; BAC08174.1; -.
 KW Complete proteome.
 SQ SEQUENCE 655 AA; 70984 MW; 36E5ABABF89A2D8 CRC64;

Query Match 45.2%; Score 1455; DB 16; Length 655;
 Best Local Similarity 47.8%; Pred. No. 1.9e-97;
 Matches 298; Conservative 108; Mismatches 201; Indels 16; Gaps 8;
 10 LKNIHTPADIRALSKDLOQLADVEGYLTHTVYISGSHFAAGTVELTVALHYEYFN 69
 20 LSELHTNVOAGLSIADLTQIAQIRKHEIYVAATGHLGPGVVELTVALVOTDLE 79
 70 VDOLWVDVGHQAYPHKILITGRKERMPITRTLGVSAPPADESEYDAFGVSHSTISAA 129
 80 KDRVWVDVGHQAYPHKILITGRKERMPITRTLGVSAPPADESEYDAFGVSHSTISAA 139
 130 LGAALIASQLRGEDEKQVAIIIGDSITGMAVEAMNNAHGDV--NAVLTINDMSISP 188
 140 LGAALIASQLRGEDEKQVAIIIGDSITGMAVEAMNNAHGDV--NAVLTINDMSISP 199
 189 GANNYLTIVLSKPYSSVRESKALAKMPSVVE---LAKTEEHVKMIVP--GTLF 242
 200 GAIRYINKRLSLQVQFPTDNLMEQFPHIIPFGENLTPEMQRLEKGRILAVRQVAVF 259
 243 EELGNNYRPGIDGHVEMLVSTLENLKL--TGPFVLTWTKKGGVAPAKD 302
 260 BELGFTYVGPVGNLRLATFQHAHTTGPVAVHAYVAGKVAIAEKQVGHNAQNP 319
 303 PD-PTKDFLPKAAPEPHPTVEFGRLCDMAADERLIGITPARESGSGVBSQKPN 361
 320 FDLVTGAKESKRP--PSYKVFGLTTLKAEVDPIVGTITAMANTGDIIDQKVPK 378
 362 RYFDVALISQHAYTLAAGQACQAKVVAIYSTFLQRGDOLIHVALON 421
 379 QYIDVGLABOAVYMAAGMAQGRPVAAIYSTFLQRAYDQIVHVCIOKLPVFCMDRA 438
 422 GLVGPDPTHAGAFDYSYRCIPNNLIMAPADENECROWLTGFP--HGPASVRYPRSGK 480
 439 GIVGADGTHQGMIDIALIRCLPNNVIMAPDDEALQEMVITGINTYTGPIALRYPRGNG 498
 481 PGAAI-DPTLALBTGKAEVRHSGRIIILAMGSVTPPAVAKOL---GATVNNR 534
 499 YGVALMEBGMEPLBEGKELRLSGEDILVAYGSMVYPAWQVAVILKHEGMSAAVYNARF 558

```

QY 535 VKPFDALVELARFTHDVFTVEENVIAGAGSALINTFLAQKYLMPVCMGLPDRFEQ 594
DB 559 AKPDTLLTFLPAQIGRWVTLIEGCGGSAVLENALRADILVFLKGVDPDIVER 618
QY 595 GSRELLIVGLDSKGLIATIEQ 617
DB 619 ASPDESKADIGLTPPMETIMO 641

RESULT 9
Q8L692 PRELIMINARY: PRT; 711 AA.
AC Q8L692;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)
DE 1-deoxy-D-xyulose 5-phosphate synthase 2 precursor,
GN DKS2.
OS Medicago truncatula (Barrel medic.).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
OC NCBI_TaxId=3880;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Root;
RA Walter M.H., Hans J., Strack D.;
RT "Two distantly-related genes encoding 1-deoxy-D-xyulose 5-phosphate
RT synthases: differential regulation in shoots and apocoteneoid-
RT accumulating mycorrhizal roots."
RL Submitted (FBS-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ430048; CAD22531.1; -.
DR InterPro; IPR005477; Dxs.
DR InterPro; IPR005476; Transketolase_C.
DR InterPro; IPR005475; Transketolase_CR.
DR InterPro; IPR005474; Transketolase_N.
DR Pfam; PF02780; transketolase_C_1.
DR Pfam; PF02779; transketolase_Pyr_1.
DR TIGRFAMs; TIGR00204; dxs; 1.
DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
KW Signal.
FT SIGNAL. 1 44 POTENTIAL.
FT CHAIN. 45 711 1-DEOXY-D-XYULOSE 5-PHOSPHATE SYNTHASE.
SQ SEQUENCE 711 AA; 76940 MW; E920A2B8E2F8C6A CRC64;

Query Match
Best Local Similarity 44.9%; Score 1444.5; DB 10; Length 711;
Matches 302; Conservative 112; Mismatches 192; Indels 27; Gaps 10;

QY 8 PLKKNITPADIRALSQOLQOLADENVGYLTHTVISGSGHFAAGLVETLVLAHYVN 67
DB 67 PLIDTVNPFVNMKNLTTEDEQLAELRADIVHSVDTGHLSSSLGVVELSVLHHVFD 126
QY 68 TPVDQLVWDVGHQAYPHKILITGRKEMPTIRTLGVSAPFAPRDESEDAFGVGHSTIS 127
DB 127 TDDDKIMVGHQAYPHKILITGRSRMHTIRKTSGLAGPKKDESVDAFGVGHSTIS 186
QY 128 AALGAVALSOLGSDKQKVAIIIGDSITGCAVYEMNNAADVANNLVILINDNMSISP- 186
DB 187 AGIGMAVAVADLLGKQKSVISVIGDGAMTAGAYEAMNNAAGFIDSLNLYILANNKQVSLPT 246
QY 187 -----PVGANNVLTLYLSSKFYSVRESKALACMPQVWEL-ARKTEEHVKGMI- 236
DB 247 ATLDPATPVGALSSTLSKIQASRKFKRLKREATKNTIKQIGQTHLVASKVDKARDFIS 306
QY 237 -VPGTLFEELGNYFGPIDGHEVMVSTLENKDL--TGYPLHVTYTKKGGYAPAKED 293
DB 307 GSGSSIFEEELGMYIYGPMGNHIEDLVNIFEKYKAMPAPGPLYIHIVTEKGGYEPALAA 366
QY 294 PLAYHGVPAFDDPKQFLPKAAPSHPHTYEVFGMLCDMAADBERLGLTPMRREGSLV 353
DB 367 ADRMGGVVFDPKTHQFKPKDST-LATYQYFADSLIKELKEDNKIKVALHAAAGGTGLN 425

```

```

QY 354 EFSQKZPNRYFVPAVLAECNAVTLAAGQACQAPVVAIVSTFLQSGDQLIHDAVQND 413
DB 426 YQKRPDPCFVQVLAECNAVTLAAGQACQAPVVAIVSTFLQSGDQLIHDAVQND 485
QY 414 MFALDRAGLVGDDGFTTANAGADYSYMRCTPNMLIAPADEVBCROMLTGRO-HHGPA 472
DB 486 VRFAMDRAGLVGDDGFTTANAGADYSYMRCTPNMLIAPADEVBCROMLTGRO-HHGPA 545
QY 473 VRFAMDRAGLVGDDGFTTANAGADYSYMRCTPNMLIAPADEVBCROMLTGRO-HHGPA 527
DB 546 FRRPRGNGIGANILPNNKGTPIELIGGRILLBSGRVAILLIGCMVOQCKKAAEMLA 605
QY 528 --TVNNRPFVPEPDALVELARFTHDVFTVEENVIAGAGSALINTFLAQKYL---MPV 582
DB 606 YVTVADARFCKPDTDLIRLAREHEILITVEGSI-GGFGSHVSQFLSLAGLDGPELKL 664
QY 583 CNIGLADRFVQSGRELLIVGLDSKGLIAT 615
DB 665 RSMMLPDRYIDHGAQPDQIDRAGLSKHLATV 637

RESULT 10
Q8L6C6 PRELIMINARY: PRT; 715 AA.
AC Q8L6C6;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)
DE 1-deoxy-D-xyulose 5-phosphate synthase.
GN DKS.
OS Stevia rebaudiana (Stevia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC Eupatoriaceae; Stevia.
OC NCBI_TaxId=55670;
RN [1]
RP SEQUENCE FROM N.A.
RC Tote N.M.L.C.;
RA "Cloning and expression of cDNAs encoding two enzymes of the MEP
RT pathway in Stevia rebaudiana Bertoni."
RL Plant Physiol. 0:0-0(0).
DR EMBL; AJ429232; CAD22155.2; -.
DR InterPro; IPR005477; Dxs.
DR InterPro; IPR005476; Transketolase_C.
DR InterPro; IPR005475; Transketolase_CR.
DR Pfam; PF02780; transketolase_C_1.
DR Pfam; PF02779; transketolase_Pyr_1.
DR TIGRFAMs; TIGR00204; dxs; 1.
DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
SQ SEQUENCE 715 AA; 76559 MW; A6C430C8FA2AFEBB CRC64;

Query Match
Best Local Similarity 44.8%; Score 1440.5; DB 10; Length 715;
Matches 303; Conservative 108; Mismatches 193; Indels 31; Gaps 12;

QY 8 PLKKNITPADIRALSQOLQOLADENVGYLTHTVISGSGHFAAGLVETLVLAHYVN 67
DB 68 PLIDTVNPFVNMKNLTTEDEQLAELRADIVHSVDTGHLSSSLGVVELSVLHHVFD 127
QY 68 TPVDQLVWDVGHQAYPHKILITGRKEMPTIRTLGVSAPFAPRDESEDAFGVGHSTIS 127
DB 127 TDDDKIMVGHQAYPHKILITGRSRMHTIRKTSGLAGPKKDESVDAFGVGHSTIS 186
QY 128 AALGAVALSOLGSDKQKVAIIIGDSITGCAVYEMNNAADVANNLVILINDNMSISP- 186
DB 128 TPDDKIMVGHQAYPHKILITGRSRMHTIRKTSGLAGPKKDESVDAFGVGHSTIS 187
QY 128 -----PVGANNVLTLYLSSKFYSVRESKALACMPQVWEL-ARKTEEHVKGMI- 236
DB 247 ATLDPATPVGALSSTLSKIQASRKFKRLKREATKNTIKQIGQTHLVASKVDKARDFIS 306
QY 237 -VPGTLFEELGNYFGPIDGHEVMVSTLENKDL--TGYPLHVTYTKKGGYAPAKED 293
DB 307 GSGSSIFEEELGMYIYGPMGNHIEDLVNIFEKYKAMPAPGPLYIHIVTEKGGYEPALAA 366
QY 187 -----PVGANNVLTLYLSSKFYSVRESKALACMPQVWEL-ARKTEEHVKGMI- 236
DB 248 ATLDPATPVGALSSTLSKIQASRKFKRLKREATKNTIKQIGQTHLVASKVDKARDFIS 307

```

QY 238 P--GILFEEGPNRYGPIIDGHVEMLVSTLENLKD--TGPPVLAHYVTKKGGYAPAEKD 293
 DB 308 ASGSTLFEELGYYIGPVDGHNVEDLVNIFEKYKSPAPGAPVLIHIVTEKGGYPPAEAA 367
 QY 294 PLAYHGVAPFD-PT-KDPLPKAASPHPTYTEVFGWLCMDMAOBERLIGTPARESG 351
 DB 368 ABRMGVAVFDPVPTKQKFTK--SPTLSYTOYFASLIKEBANDKYLAIHAAAGGATG 424
 QY 352 LVEFSQKPNRYFDVAIAQAVTLAAGACQAGKPVVAIYSTPTORGTDLIHVALON 411
 DB 425 LNYFQKCEPCERCFDVGIAQHAVTFAGLATEGLKPFCAIYSSFLORGDYQVHVHDLOK 484
 QY 412 LMLFLDAGLVGPGPPTHAGAPDYSYKCI PNMLIMAPADENECROWLTGFGH-HNEP 470
 DB 485 LVEFPMADAGLVGADGPHGAPDITVACLPNMVWAPADAEIMHVAAPAAALIDRRP 544
 QY 471 ASVRYPRGKPGAAIDPTLTALA--LEIGKARVHHGSRILAIAMGSMVTPAVEAKOL--- 525
 DB 545 SCRRPFRGNGIGAPLEPNKNGIPIEVGKGRILLEGTRVAILGVSTIGVGCAGASILQAH 604
 QY 526 --GATVNMRFVPPDQALVLELARTHDVPTVEENVVAGGASINTLQOKVL--M 580
 DB 605 NVSATVADARFCPPDITGLIRLANEHEVLLTYEBGSI-GGFGSHVAFSLINGLIDKTL 663
 QY 581 PVCNIGLPDRFVQSGREELSLVGHDSKGIATI 615
 DB 664 KLRAMTLPRKIDHGAPODQLEBTGLSSHGCSL 698

RESULT 11

Q8K213 PRELIMINARY; PRT; 630 AA.
 AC Q8K213;
 DT 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE deoxyxylulose-5-phosphate synthase.
 GN DXS.
 OS uncultured proteobacterium.
 OC Bacteria; Proteobacteria; environmental samples.
 NC NCBT_TaxID=153809;

RP SEQUENCE FROM N.A.
 RX MEDLINE=21822632; PubMed=11832943;
 RA Beja O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M.,
 Hamada T., Eisen J.A., Fraser C.M., DeLong E.F.,
 RT "Unsuspected diversity among marine aerobic anoxygenic phototrophs."
 RL Nature 415:630-633(2002).
 DR EMBL; AE008921; AAM48660.1; --
 DR InterPro; IPR005477; Dxs.
 DR InterPro; IPR005476; Transketolase C.
 DR InterPro; IPR005475; Transketolase C.
 DR InterPro; IPR005474; Transketolase N.
 DR Pfam; PF02780; transketolase_C_1.
 DR Pfam; PF02779; transket_ptyr_1.
 DR TIGRfams; TIGR00204; dxs_1.
 DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
 DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
 SQ SEQUENCE 630 AA; 65840 MW; 26789DADF48111C97 CRC64;

Query Match 44.7%; Score 1437.5; DB 2; Length 630;
 Best Local Similarity 47.6%; Pred. No. 3 4e-96;
 Matches 302; Conservative 100; Mismatches 195; Indels 37; Gaps 9;

QY 7 YPLKNHTPADIRALSKDQLQDLADEVRGYLTHVTSIGGHPAAGLGTVELTVALHYVF 66
 DB 2 HSLIIEITPSDKITSLDQQLADDELRVDIENASRFGHSGSLGVELTVALHAFV 61
 QY 67 NTPVQDLVWDVQHAQVPHKILNGRKEKPTIRLIGSVSAFPAADSEVYAPGVGSSSTSI 126
 DB 62 DTPKDKLMDVGVHQCPYHVKVITGRRAKGTILROGGSLFTKRSSEVPYPPAAHSSSTSI 121
 QY 127 SAALGMAIASQLRGEDKKNVAIIGDSITGGAYEAMNAGDVNANLLVILNDMSISIP 186

DB 122 SAHGFVARDLGGDTGDALIVIDGSLASGMAIYBALNAGABRRFPVILINDMEKSTAP 181
 QY 187 PVGAMNNYITVYLSKPYSSVREESKALAXMPS-----VWELARKTEEHYXG 234
 DB 182 PVGAMSKMMSGLAGI-----ALAQINFGQDIETVLDPQPRDHRARRARELVATG 229
 QY 235 MIV-PGILFEEGPNRYGPIIDGHVEMLVSTLENLKD--LTGPVLAHYVTKKGGYAPAEK 292
 DB 220 AVASGQITFEELGYYIGPVDGHNVEDLVNIFEKYKSPAPGAPVLIHIVTEKGGYPPAE 289
 QY 293 PLAYHGVAPFDPTKAPSPHPTYTEVFGWLCMDMAOBERLIGTPARESG 352
 DB 290 SADKTHGVAKDVASGAMSGANA-PSTYTFEGTTLKALAKOSKIVGITAAHPSGGL 348
 QY 353 VEFQKPNRYFDVAIAQAVTLAAGACQAGKPVVAIYSTPTORGTDLIHVALON 412
 DB 349 DIFAPRPFADHEDVIGIAEQHGVTFAGAGLAPFCIAIYSTPTORGTDLIHVALON 408
 QY 413 DMLFLDAGLVGPGPPTHAGAPDYSYKCI PNMLIMAPADENECROWLTGFGH-HNEP 471
 DB 409 PVRFADAGLVGADGPHGAPDITVACLPNMVWAPADAEIMHVAAPAAALIDRRP 544
 QY 472 SVRYPRGKPGAAIDPTLTALA--LEIGKARVHHGSRILAIAMGSMVTPAVEA---GKOLG 526
 DB 469 AFRYPRGTGVSIPQEGELLQIGKRIYRGAELALLSFGHLAEALKAADLISAQGV 528
 QY 527 ATVNMRFVPPDQALVLELARTHDVPTVEENVVAGGASINTLQOKVL--M 583
 DB 529 ATVADARFAPKPLDHALISKLAKTRVLLITIOGA-QGFGAMVLAHYVADTGLDGLPLAVR 587
 QY 584 NIGLPDRFVQSGREELSLVGHDSKGIATI 617
 DB 586 SMTLPRFDITDQAPADPAMVADAGLATDIAATLQ 621

RESULT 12

Q9FSG2 PRELIMINARY; PRT; 709 AA.
 AC Q9FSG2;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE 1-D-desoxyxylulose 5-phosphate synthase (Dxs).
 GN DXS.
 OS Narcissus pseudonarcissus (Daffodil).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
 OC Narcissus.
 NC NCBT_TaxID=39639;

RP SEQUENCE FROM N.A.
 RA Schaub P., Beyer P., Al-Babli S.,
 RT "A cDNA encoding 1-D-desoxyxylulose 5-phosphate synthase (Dxs) from
 RL Narcissus pseudonarcissus L.";
 DR Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ279019; CAC08458.1; --
 DR InterPro; IPR005477; Dxs.
 DR InterPro; IPR005476; Transketolase C.
 DR InterPro; IPR005475; Transketolase C.
 DR InterPro; IPR005474; Transketolase N.
 DR Pfam; PF02780; transketolase_C_1.
 DR Pfam; PF02779; transket_ptyr_1.
 DR TIGRfams; TIGR00204; dxs_1.
 DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
 DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
 SQ SEQUENCE 709 AA; 75747 MW; 1FBA362A0DEDC6B3 CRC64;

Query Match 44.6%; Score 1435.5; DB 10; Length 709;
 Best Local Similarity 47.6%; Pred. No. 5.8e-96;
 Matches 302; Conservative 107; Mismatches 195; Indels 31; Gaps 11;

QY 8 PLKNHTPADIRALSKDQLQDLADEVRGYLTHVTSIGGHPAAGLGTVELTVALHYVF 67
 DB 6 PLKNHTPADIRALSKDQLQDLADEVRGYLTHVTSIGGHPAAGLGTVELTVALHYVF 67

DB 65 PLDTITVPARKKNSLTDLEOLASRLDIHVSSTGGHLSASLGVNDLTVLHVVD 124

QY 68 TPVDQVWDVGHQAYPHKILITGRKRMPTITRTIGVSAFPADESEYDAFGVSHSTIS 127

DB 125 TDDDKIMVDVGHQAYPHKILITGRSRMHTLRQTSGLAGFPKDESDVDAFGVSHSTIS 184

QY 128 AALGMAIASQLRGEDKQKVAIIGDSTITGMAIYEMNAGDVANALVILNDNISIP- 186

DB 185 AGLGMAVAGRLDQKSGNHSVIGDGMVTAQAYEMNNVNTGQVILVILNDNQVSLPT 244

QY 187 -----PVGAMNNYLTKYLSSKFYSVRESKALAKM-PSVELARKTEEHVKMIV 237

DB 245 AALNGPAPVGAHSGALALQASAKPRLREAKICITQIGGQAEVAAKVDYARGNIS 304

QY 238 P--GTLFEEELGPNYFPGIDGHDVEMLVSTLENKDL--TGPFVLAHVTTKKGQYAPAEKD 293

DB 305 ASGASLFEELGLYIYIGVDGSHVEDLVAFKKYKAMPSPGPAVIAIVTEKKGQYPAEAA 364

QY 294 PLAYHGVPAHPDT--KDFLPKAPSPHPTYTEVGRMLCDMAAGDELLGITPARBEGSG 351

DB 365 AAKRGVAVKFDKTKGQFKTK--PPTLSTYGFABSLYKAEVDEKIVAIHAAMGGGTG 421

QY 352 LVEFSQKFNRYFDVAIAEOHVAVTLAAGQACQAKPVVAIYSTLQRGYDQILHDAVALON 411

DB 422 LNFQKKEFDRCPDVGAIEQHAHTFAAGLATEGLKPCFCAIYSSFLQRGYDQVHDVLOK 481

QY 412 LDMFLALDPRAGVGPDPHTHAGAFDYSYRCIPNNLIMAPDEBCROMLTTGQ--IHGP 470

DB 482 LVPFRLDAGLVADGPHCAFDVTAACLPNNIWMASDEALMNNVATAAIDRRP 541

QY 471 ASVRYPRGKPGAAI--DPTLTALSIGKAEVRHSGRIALIAMGSMTPPAVEAG-----K 523

DB 542 SCFRPFRNGVAVLPSDYKGTPLIEKGRIIMBQDKVALIGVSIYVSCLEKAASTLER 601

QY 524 QLGATVNNKRFKPFDOALVLELARTHDVFTVVEENVIAAGAGSINTFLQAOVL---M 580

DB 602 GSATVADARCFKPLDSELIRLVNEHELITVEEGSI--GFPASHVSHLSLGLDGL 660

QY 581 PVCNIGLPRFVQGSREBELSLVGLDSKGLIATI 615

DB 661 KLRSMVLPRIYIDHGAHPKQIEAGLSKHNGTV 695

RESULT 13

Q9SP65 PRELIMINARY: PRT; 713 AA.

AC Q9SP65: 01-MAY-2000 (TREMREL. 13, Created)

DT 01-OCT-2000 (TREMREL. 15, last sequence update)

DT 01-OCT-2002 (TREMREL. 22, last annotation update)

DE 1-deoxy-D-xylose-5-phosphate synthase.

GN DXS1.

OS Artemisia annua (Sweet wormwood).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; Campanulids; Asterales; Asteraceae; Asteroideae;

OC Arctemiaceae; Artemisia.

OC NCBI_TaxID=35608;

OX NCBI_TaxID=35608;

RN RN

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Yugoslavica; TISSUE=Root;

RA Wobbe K.X., Souret F.F., Shore K.A., Weathers P.J.;

RT "Artemisia annua D-1-deoxyxylose-5-phosphate synthase (dxs) mRNA.";

RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.

DR BMBJ; A182286; AAD56390.2; -.

DR InterPro; IPR005477; Dxs.

DR InterPro; IPR005476; Transketolase_C.

DR InterPro; IPR005475; Transketolase_CR.

DR InterPro; IPR005474; Transketolase_N.

DR Pfam; PF02780; transketolase_C_1.

DR Pfam; PF02779; transket_pyr_1.

DR TIGRFAMs; TIGR00204; dxs_1.

DR PROSITE; PS00801; TRANSKETOLASE_1; 1.

SO SEQUENCE 713 AA; 76669 MW; 4D3FF36512C2377A CRC64;

Query Match 44.6%; Score 1435.5; DB 10; Length 713;

Best Local Similarity 47.5%; Pred. No. 5.9e-96;

Matches 303; Conservative 111; Mismatches 187; Indels 37; Gaps 13;

QY 8 PLNKHITPADIRALSDQLOQLADEVRGYLITHTVSIISGCFAGLCTVELTVAHVFN 67

DB 68 PLDTITVPARKKNSLTDLEOLASRLDIHVSSTGGHLSASLGVNDLTVLHVVD 127

QY 68 TPVDQVWDVGHQAYPHKILITGRKRMPTITRTIGVSAFPADESEYDAFGVSHSTIS 127

DB 125 TDDDKIMVDVGHQAYPHKILITGRSRMHTLRQTSGLAGFPKDESDVDAFGVSHSTIS 184

QY 128 AALGMAIASQLRGEDKQKVAIIGDSTITGMAIYEMNAGDVANALVILNDNISIP- 186

DB 185 AGLGMAVAGRLDQKSGNHSVIGDGMVTAQAYEMNNVNTGQVILVILNDNQVSLPT 244

QY 187 -----PVGAMNNYLTKYLSSKFYSVRESKALAKMPS-VWELARKTEEHVKMIV- 236

DB 248 ANDGPPIPPVGAHSGALALQASAKPRLREAVAGVTKQIGPHEHLAAKVDYARGNIS 307

QY 237 -VPGLFEEELGPNYFPGIDGHDVEMLVSTLENKDL--LTPGVLAHVTTKKGQYAPAEKD 293

DB 308 GSGSTLEELGLYIYIGVDGSHSIDLVAILKEVSTRTTGTVLHVITEKRGYPAEKA 367

QY 294 PLAYHGVPAHPDT--KDFLPKAPSPHPTYTEVGRMLCDMAAGDELLGITPARBEGSG 351

DB 368 ADKTHGVGRKDPATGKPFKSA--PTQSTYTFABSLYKAEVDEKIVAIHAAMGGGTG 424

QY 352 LVEFSQKFNRYFDVAIAEOHVAVTLAAGQACQAKPVVAIYSTLQRGYDQILHDAVALON 411

DB 425 LNFQKKEFDRCPDVGAIEQHAHTFAAGLATEGLKPCFCAIYSSFLQRGYDQVHDVLOK 484

QY 412 LDMFLALDPRAGVGPDPHTHAGAFDYSYRCIPNNLIMAPDEBCROMLTTGQ--IHGP 470

DB 485 LVPFRLDAGLVADGPHCAFDVTAACLPNNIWMASDEALMNNVATAAIDRRP 544

QY 471 ASVRYPRGKPGAAI--DPTLTALSIGKAEVRHSGRIALIAMGSMTPPAVEAG-----K 523

DB 545 SCFRPFRNGVAVLPSDYKGTPLIEKGRIIMBQDKVALIGVSIYVSCLEKAASTLER 604

QY 524 QLGATVNNKRFKPFDOALVLELARTHDVFTVVEENVIAAGAGSINTFLQAOVL--- 579

DB 605 GLNITVADARCFKPLDSELIRLVNEHELITVEEGSI--GFPASHVSHLSLGLDGL 663

QY 580 --MPVCNIGLPRFVQGSREBELSLVGLDSKGLIATI 615

DB 664 KMRPLV---LPDKYIDHGAHPADQLAERGLTPSHIATV 698

RESULT 14

Q8GAA0 PRELIMINARY: PRT; 636 AA.

AC Q8GAA0: 01-MAR-2003 (TREMREL. 23, Created)

DT 01-MAR-2003 (TREMREL. 23, last sequence update)

DT 01-MAR-2003 (TREMREL. 23, last annotation update)

DE 1-deoxy-D-xylose 5-phosphate synthase (EC 4.1.3.37).

GN DXS.

OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).

OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.

OC NCBI_TaxID=1140;

OX NCBI_TaxID=1140;

RN RN

RP SEQUENCE FROM N.A.

RC MEDLINE=20026086; Pubmed=10556522;

RA Miller B., Heuser T., Zimer W.;

RT "A Synecococcus leopoliensis SAUG 1402-1 operon harboring the 1-deoxyxylose 5-phosphate synthase gene and two additional open reading frames is functionally involved in the dimethylallyl diphosphate synthesis.";

RL FBS Lett. 460:485-490(1999).

RN [2]

RP SEQUENCE FROM N.A.
RA Holman C.K., Sandoval P., Chen Y., Socias T., Mohler B.J.,
RA McMurtry S., Gonzalez A., Salinas I., Golden S.S., Youderian P.,
RT "Synechococcus elongatus PCC 7942 cosmid 389."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RM EMBL: X04616; CDS5546.1; ..
KM Lyase.
SQ SEQUENCE 636 AA; 69328 MW; 51B38C3A5A7A3D CRC64;
Query Match 44.6%; Score 1433; DB 2; Length 636;
Best Local Similarity 48.1%; Pred. No. 7.4e-96;
Matches 300; Conservative 99; Mismatches 209; Indels 16; Gaps 8;
QY 10 LKNHTPADIRALSKDQIQQLADREVRGILHTVTSISGHPAAGTVELTVALHYVNTF 69
DB 3 LSEITHPQAGLSVAQIQEQIGHQIRKHHQITVAAGHLPAGVVELTVALYQTLDD 62
QY 70 VDQVWDVGHQAYPHKILITGRKEMPTIRTLGVSAPFAPDESEYDAFGVSHSTISAA 129
DB 63 RDKVWVDVGHQAYPHKILITGRKEMPTIRTLGVSAPFAPDESEYDAFGVSHSTISAA 122
QY 130 LGMALASQSGEDKKVAIIGDGSITGMAYEAMNAGDV-NANLVIANDMSISPPV 188
DB 123 LGMALASQSGEDKKVAIIGDGSITGMALAEINHAHLEKTRLLVANDMSISPPV 182
QY 189 GAMNNYITKVLSSKPYSSVREBSKKALAKMPSVWEAKTEEHK-GM-----IVPGLF 242
DB 183 GALSTYIAKLIVSEPMQLITDGLTQGMQILPFGAGALTQGFPEVKEGKRLSYKIGAVF 242
QY 243 BELGFNYFPIGDHVMVLSTLENLXDLTPVFLAVVTKKGKGYAPAEKDPPLAHGVPA 302
DB 243 BELGFNYFPIGDHVMVLSTLENLXDLTPVFLAVVTKKGKGYAPAEKDPPLAHGVPA 302
QY 303 PD-PTKQPLPKAASPHPTTYTEVGRMLCDMAODERLGITTAMRBSGGLVESQKPEP 361
DB 303 PD-PTKQPLPKAASPHPTTYTEVGRMLCDMAODERLGITTAMRBSGGLVESQKPEP 361
QY 362 RYFPAVIAEQAVALTAAGACOGAPVVAIYSTFLOGGYDQILHDVALONLMLFALDRA 421
DB 362 RYFPAVIAEQAVALTAAGACOGAPVVAIYSTFLOGGYDQILHDVALONLMLFALDRA 421
QY 422 GLVDEGDPHAGAFDYSTWRCIPNMLIMAPADENECROMLTTFQSHHAPSVRYPRKGP 481
DB 422 GIVADGDPHAGAFDYSTWRCIPNMLIMAPADENECROMLTTFQSHHAPSVRYPRKGP 481
QY 482 GAAL-DPTLTALIKGAEVRHHSRIALLANGSVTPAVEAGKQ-----GATVNMRFV 535
DB 482 GVPLEPBGWESLPFGKAEQLAQGDLLMLATGSKVTPALQTAELNHEGISAIVINRFA 541
QY 536 KFPDQALVLELARTHDVFEVTEVENVIAGAGSAINTEFLAQKVLMPVCNIGLPDRFVEOG 595
DB 542 KFLDEBELIVPLARQIGKRVVTFEEGCLPGFGSALIMESIQANDLQVPLPIGVPDLIVEHA 601
QY 596 SREELSLVIGDSKGLATI-EQF 618
DB 602 SPDESKQELGLTPROMADRILEKF 625
RESULT 15
OSMAYL PRELIMINARY; PRT; 722 AA.
AC OSMAYL
DT 01-MAR-2002 (TREMBLrel. 20; Created)
DT 01-MAR-2002 (TREMBLrel. 20; Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 23; Last annotation update)
DE Deoxy-D-xylose-5-phosphate synthase.
GN DKS.
OS Morinda citrifolia (Indian mulberry).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
OC Asteridae; Lamiales; Gentianales; Rubiaceae; Rubioideae; Morinda;
OC Morinda.
OK NCBI_TaxID=43522;

RN [1]
RA SEQUENCE FROM N.A.
RA Han Y.S., Roytcharuk S., Verberne M.C., Heijden, Rvd, Verpoorte R.,
RT "Cloning and characterization of a cDNA encoding 1-deoxy-D-xylose 5-
RT phosphate synthase from Morinda citrifolia."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RM EMBL: AF43590; AAJ32062.1; ..
KM InterPro: IPR005477; Dxs.
DR InterPro: IPR005476; Transketolase_C.
DR InterPro: IPR005475; Transketolase_CR.
DR InterPro: IPR005474; Transketolase_N.
DR Pfam: PF02780; transketolase_C; 1.
DR Pfam: PF02779; transketolase_Pyr; 1.
DR TIGRfam: TIGR00204; dxs; 1.
DR PROSITE: PS00801; TRANSKETOLASE_1; 1.
DR PROSITE: PS00802; TRANSKETOLASE_2; 1.
SQ SEQUENCE 722 AA; 77970 MW; 58231942EAF9C767 CRC64;
Query Match 44.5%; Score 1430.5; DB 10; Length 722;
Best Local Similarity 47.1%; Pred. No. 1.4e-95;
Matches 302; Conservative 104; Mismatches 198; Indels 37; Gaps 12;
QY 8 PLKNHTPADIRALSK-----DQLQQLADREVRGILHTVTSISGHPAAGTVELTVA 61
DB 72 PLDTINYPPIHMKNSLSTKSHYLIQELKQLADRELSDIIFVSKTGGHLSGLVVELTVA 131
QY 62 LHYVNTPVQVWDVGHQAYPHKILITGRKEMPTIRTLGVSAPFAPDESEYDAFGVGH 121
DB 132 LHYVNTPVQVWDVGHQAYPHKILITGRKEMPTIRTLGVSAPFAPDESEYDAFGVGH 121
QY 122 SSTISIAALGMAIASQLRGEKKVAIIGDGSITGMAYEAMNAGDVANLVIANDND 181
DB 192 SSTISIAAGLGMAYGRLLGKNNNVISVIGGAVTAAQAYEAMNAGDLNLIIVINDNK 251
QY 182 MSISP-----PYGAMNNYITKVLSSKPYSSVREBSKKALAKM-PSVWEAKTEEH 231
DB 252 QVSLPTGLDPAATPGALSSALTTLQASPKFQLEBAASITKQIGPQAEVAANKDY 311
QY 232 VKGMATVP--GLTFEELGPNVFGPIGDHVMVLSTLENLXDL--TGFVFLAVVTKKGKGY 287
DB 312 ARGILASGTFEELGLTYIGVDGHNIEDLITTDKATMTPAPGVLIHYTEKKGKGY 371
QY 288 APAEKDPPLAHGVPAFPT--KQLEPKAASPHPTTYTEVGRMLCDMAODERLGITPA 345
DB 372 PPAEAAADRHHGVKFDVVASGQFKSK--SPTLSYQYFAELILKEALIDNKIVAIHAA 428
QY 346 MREGSGIVSESQKPEPPEVPAVIAEQAVALTAAGACOGAPVVAIYSTFLOGGYDQILH 405
DB 429 MGGGTGLNTFQKLFPERCEPVGIAEGHVAVFAAGLITBGLKPCALYSTFLOGGYDQVH 488
QY 406 DVALONLMLFALDRAGLVGPDPPTAGAFDYSTWRCIPNMLIMAPADENECROMLTTFG 465
DB 489 DVDQLQKLPVFPANDRAGLVGADGPTGCAFDVAYMCLSNMIWAPADEALNEMVATA 548
QY 466 Q-HHGRASVRYPRKGPAGALIDPTL--TALETGKAEVRHHSRIALLANGSVTPAVEAG 522
DB 549 TIDRPPCFRFPFGKNGIGAKLPDKNKENTLIGKGRILTEGSSVAILIGYALVQCCGAA 608
QY 523 KQD-----GATVNMRFVFPDQALVLELARTHDVFEVTEVENVIAGAGSAINTEFLAQK 577
DB 609 EMUKLNRIKPTIADAFRCRDLDELKRLAKHEHILLIYVEGSI--GGPGSHVHPLSLSG 667
QY 578 VL--MPVCNIGLPDRFVEOGSREELSLVIGDSKGLATI 615
DB 668 LLDGPIKLRSMVLPDRYIDHGSPPADQIEAGLSRHICATV 708

Search completed: January 29, 2004, 15:54:02
Job time : 49.2142 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:40:33 ; Search time 36.0149 Seconds
(without alignments)
1736.455 Million cell updates/sec

Title: US-09-941-947a-8

Sequence: I MKGICILGATGSGVSTLDV.....IQADPDAREVADIKITLVA 394

Scoring table: BLASTSUM62
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1985	100.0	394	23 ABG61582
2	1985	100.0	394	23 AA021877
3	1985	100.0	394	23 AA021887
4	1215.5	61.2	396	23 AA021884
5	1052.5	53.0	394	23 AA021886
6	1045	52.6	398	23 AA021878
7	1045	52.6	398	23 AA021885
8	1043.5	52.6	411	24 ABP79699
9	1043.5	52.6	411	24 ABP80486

10	1042	52.5	400	23 AA021889
11	1007	50.7	397	23 AA021877
12	1007	50.7	397	23 AA021887
13	1007	50.7	397	23 AB09436
14	999.5	50.4	388	23 AA021879
15	999.5	50.4	388	23 AA021883
16	999	50.3	398	20 AA021883
17	994.5	50.1	386	23 AA021875
18	892.5	45.0	394	23 AA021881
19	892.5	45.0	394	23 AA021888
20	886	44.6	402	23 AA021880
21	886	44.6	402	23 AA021890
22	863	43.5	295	21 AA021890
23	840	42.3	295	21 AA021890
24	818.5	41.2	399	21 AA021890
25	818.5	41.2	406	21 AA021890
26	818.5	41.2	406	21 AA021890
27	818.5	41.2	420	21 AA021890
28	818.5	41.2	476	21 AA021890
29	818.5	41.2	477	22 AA021890
30	818.5	41.2	477	23 AA021890
31	818.5	41.2	477	23 AA021890
32	818.5	41.2	477	23 AA021890
33	817.5	41.2	477	21 AA021890
34	816.5	41.1	380	23 AA021890
35	814.5	41.0	477	22 AA021890
36	810.5	40.8	477	21 AA021890
37	802.5	40.4	472	21 AA021890
38	802.5	40.4	472	21 AA021890
39	801	40.4	250	21 AA021890
40	798.5	40.2	388	23 AA021890
41	798.5	40.2	388	23 AA021890
42	795.5	40.1	477	21 AA021890
43	788.5	39.7	475	21 AA021890
44	778	39.2	394	20 AA021890
45	754.5	38.0	499	21 AA021890

ALIGNMENTS

RESULT 1
ABG61582
ID ABG61582 standard; Protein; 394 AA.
XX
AC ABG61582;
XX
DT 27-AUG-2002 (first entry)
XX
DE High growth methanotrophic bacterial strain polypeptide #32.
XX
KW High growth methanotrophic bacterial strain; C1 carbon substrate; enzyme;
KW methanol; methanol; Embden-Meyerhof carbon flux pathway; 16S RNA;
KW pyrophosphate dependent phosphotransferase; nitrogen-containing compound;
KW ammonia; nitrate; nitrite; nitrogen; oxygen; landfill;
KW ammonia-containing environment; waste water treatment system; isoprenoid;
KW nitrous oxide; terpenoid; animal feed; carotenoid; exopolysaccharide.
XX
OS Methylobionas 16a.
XX
PN MO200220728-A2.
XX
PD 14-MAR-2002.
XX
PE 28-AUG-2001; 2001MO-US26827.
XX
PR 01-SEP-2000; 2000US-229658P.
XX
PI (DUPC) DU POINT DE MEMOIRS & CO E. I.
XX
PI Koffas M, Odum JM, Schenzle A;
XX
DR WPI, 2002-452200/48.

DR N-PSDB; ABK83261.

XX New high growth methanotrophic bacterial strain, useful for producing
PT single cell proteins, grows on a C1 carbon substrate, and comprises a
PT functional gene encoding in Embden-Meyerhof carbon pathway

PS Claim 11; Page 141-142; 157pp; English.

XX The invention relates to a high growth methanotrophic bacterial strain,
CC which grows on a C1 carbon substrate e.g. methane and methanol, and
CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a
CC gene coding a pyrophosphate dependent phosphotransferase enzyme or a 16s
CC RNA. The bacterial strain is useful for the production of single cell
CC protein and for the biotransformation of a nitrogen-containing compound,
CC e.g. ammonia, nitrate or nitrogen. It is also useful for the
CC production of a feed product comprising a protein, carbohydrates and a
CC pigment and for reducing oxygen demand, for removing nitrates and
CC nitrites in methane-containing environments such as landfills, waste
CC water treatment systems or anywhere that methane, oxygen and nitrates are
CC present. The bacterial strain of the invention can be used as a
CC denitrifying agent for the conversion of nitrate or nitrite to nitrous
CC oxide with methane or methanol as a carbon source. It is also used in the
CC production of biomass including proteins, carbohydrates and a wide
CC variety of pigments (particularly for isoprenoid pigments for the
CC purpose of generating animal feeds), in production of terpenoid and
CC carotenoid compounds, useful as pigments and as monomers in polymeric
CC materials and in production of exopolysaccharides at high levels.
CC Sequences ABG61551-ABG61590 represent high growth methanotrophic
CC bacterial strain proteins of the invention.

SO Sequence 394 AA;

Query Match 100.0%; Score 1985; DB 23; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.8e-185;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGICILGATGSGIVSTLTVVARRHPKXQVVALTANNGNDALYEOCLAHPEYAVVMES 60
DB 1 MKGICILGATGSGIVSTLTVVARRHPKXQVVALTANNGNDALYEOCLAHPEYAVVMES 60
QY 61 KVAEFKORIAASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLLPTLAARKAKGT 120
DB 61 KVAEFKORIAASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLLPTLAARKAKGT 120
QY 121 VILANKKALVMSGOIFMOAVSDSGAVLLPDISSENAIFOCMPAGYTPGHTAKOARRILLT 180
DB 121 VILANKKALVMSGOIFMOAVSDSGAVLLPDISSENAIFOCMPAGYTPGHTAKOARRILLT 180
QY 181 ASGGPFRRTPIETLSSTVTPDOAVAHPKMDGKRTSVSATMANKGELIEACLLFNNMSPD 240
DB 181 ASGGPFRRTPIETLSSTVTPDOAVAHPKMDGKRTSVSATMANKGELIEACLLFNNMSPD 240
QY 241 QIEVVHPOSIHSMVDYVDGSVLAQMGNDPMDRTPIAHAAWAMPFRPDSGVAALDIFEVGH 300
DB 241 QIEVVHPOSIHSMVDYVDGSVLAQMGNDPMDRTPIAHAAWAMPFRPDSGVAALDIFEVGH 300
QY 301 MPEKEDLRFPCRLAYEAIKSGIMPTVLANAEIYEAFLNEBEVKTDTIAVIERESM 360
DB 301 MPEKEDLRFPCRLAYEAIKSGIMPTVLANAEIYEAFLNEBEVKTDTIAVIERESM 360
QY 361 AQFKPDAGSLVLTQADODARAVARDIITKTLVA 394
DB 361 AQFKPDAGSLVLTQADODARAVARDIITKTLVA 394

RESULT 2

AAE22302
ID AAE22302 standard; Protein; 394 AA.

AC AAE22302;

DT 25-JUL-2002 (first entry)

XX

DE Methylomonas 16a sp. D-1-deoxyxyjulose-5-phosphate reductoisomerase.

XX Carotenoid; isopentenyl pyrophosphate; antheraxanthin; diet;
KM anti-oxidant; steroid; flavour; fragrance; electro-optic application;
KM aquaculture; enzyme; D-1-deoxyxyjulose-5-phosphate reductoisomerase; Dxr.
XX Methylomonas 16a sp.

OS WO200218617-A2.

FN 07-MAR-2002.

XX 04-SEP-2001; 2001WO-US27420.

XX 01-SEP-2000; 2000US-229858P.

XX 01-SEP-2000; 2000US-229907P.

XX (DUPO) DU PONT DE NEMOURS & CO B. I.

XX Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;
P1 Odum JM, Plocatagis SK, Rouviere PE;

XX WPI; 2002-351711/38.

XX N-PSDB; AAD35500.

DR Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by
XX using microorganisms having a nucleic acid molecule encoding enzymes in
XX the carotenoid biosynthetic pathway and which metabolize single carbon
XX substrates

PS Claim 42; Page 112-114; 156pp; English.

CC The invention relates to a method for producing carotenoid compounds.
CC The method comprises a transformed metabolising host cell, comprising
CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule
CC encoding an enzyme in the carotenoid biosynthetic pathway, under the
CC control of regulatory sequences, and contacting the host cell with carbon
CC substrate to produce a carotenoid compound. The method is useful for
CC producing carotenoid compounds such as antheraxanthin and astaxanthin, by
CC using microorganisms having a nucleic acid molecule encoding enzymes in
CC the carotenoid biosynthetic pathway and which metabolize single carbon
CC substrates. The carotenoids have potent anti-oxidant properties useful in
CC diet, and aquaculture elements. The carotenoids are also useful as
CC intermediates in the synthesis of steroids flavours and fragrances and
CC compounds for potential electro-optic applications. The present sequence
CC is Methylomonas 16a sp. D-1-deoxyxyjulose-5-phosphate reductoisomerase
CC (Dxr) enzyme used in the invention.

SO Sequence 394 AA;

Query Match 100.0%; Score 1985; DB 23; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.8e-185;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGICILGATGSGIVSTLTVVARRHPKXQVVALTANNGNDALYEOCLAHPEYAVVMES 60
DB 1 MKGICILGATGSGIVSTLTVVARRHPKXQVVALTANNGNDALYEOCLAHPEYAVVMES 60
QY 61 KVAEFKORIAASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLLPTLAARKAKGT 120
DB 61 KVAEFKORIAASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLLPTLAARKAKGT 120
QY 121 VILANKKALVMSGOIFMOAVSDSGAVLLPDISSENAIFOCMPAGYTPGHTAKOARRILLT 180
DB 121 VILANKKALVMSGOIFMOAVSDSGAVLLPDISSENAIFOCMPAGYTPGHTAKOARRILLT 180
QY 181 ASGGPFRRTPIETLSSTVTPDOAVAHPKMDGKRTSVSATMANKGELIEACLLFNNMSPD 240
DB 181 ASGGPFRRTPIETLSSTVTPDOAVAHPKMDGKRTSVSATMANKGELIEACLLFNNMSPD 240
QY 241 QIEVVHPOSIHSMVDYVDGSVLAQMGNDPMDRTPIAHAAWAMPFRPDSGVAALDIFEVGH 300
DB 241 QIEVVHPOSIHSMVDYVDGSVLAQMGNDPMDRTPIAHAAWAMPFRPDSGVAALDIFEVGH 300

QY 301 MDEKPDLRPPCLRLAYEALIKSGIMPTVLANANEIAYEALNEBEVKETDIAYIIERSM 360
 DB 301 MDEKPDLRPPCLRLAYEALIKSGIMPTVLANANEIAYEALNEBEVKETDIAYIIERSM 360
 QY 361 AQFKPDDAGSLLEVLQADQDAREVARDIIKTIVA 394
 DB 361 AQFKPDDAGSLLEVLQADQDAREVARDIIKTIVA 394

RESULT 3
 AAU80326
 ID AAU80326 standard; Protein; 394 AA.
 XX AAU80326;
 AC
 XX 15-JUL-2002 (first entry)
 DT
 XX Methylomonas 16a ORF2 dtx protein sequence.
 DE
 XX Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;
 KW keratenoid; pigment; flavour; fragrance; open reading frame 2; ORF2;
 KM dtx; 1-deoxyxylulose-5-phosphate reductoisomerase enzyme.
 XX
 OS Methylomonas sp.
 XX
 PN MO200220733-A2.
 XX
 PD 14-MAR-2002.
 XX
 PF 29-AUG-2001; 2001WO-US26852.
 XX
 PR 01-SEP-2000; 2000US-229907P.
 XX
 XX (DUFO) DU PONT DE NEMOURS & CO E. I.
 XX
 PI Cheng Q, Koffas M, Norton KC, Odom JK, Picataggio SK, Rouviere PE;
 PI Schenzle A, Tomb J;
 XX
 DR WPI: 2002-383051/41.
 DR N-PSDB; ABR50082.
 XX
 PT Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,
 PT isolated from Methylomonas 16a, useful for the production of isoprenoid
 PT compounds -
 XX
 PS Claim 4; Page 68-70; 84pp; English.
 XX
 CC The present invention relates to a new nucleic acid molecule encoding
 CC an isoprenoid biosynthetic enzyme isolated from Methylomonas 16a.
 CC The invention is useful for obtaining a nucleic acid molecule
 CC encoding an isoprenoid compound biosynthetic enzyme, and for the
 CC microbial production of isoprenoid compounds. The molecules of the
 CC invention are also useful for regulating isoprenoid biosynthesis in an
 CC organism and for producing recombinant organisms for producing various
 CC isoprenoid compounds. The nucleic acid is also useful for feed additive,
 CC for the production of keratenoids and their derivatives, isoprenoid
 CC intermediates, and as pure products useful as pigments, flavour and
 CC fragrances. The present amino acid sequence represents the Methylomonas
 CC 16a open reading frame 2 (ORF2) dtx (1-deoxyxylulose-5-phosphate
 CC reductoisomerase enzyme) protein of the invention, as described above.
 XX
 XX
 SQ Sequence 394 AA;

Query Match 100.0%; Score 1985; DB 23; Length 394;
 Best Local Similarity 100.0%; Pred. No. 1.8e-185;
 Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGICITGATGSGTGTGTVVVARHPKQYVALTANGNDALYECIAHHPYAAVWMS 60
 DB 1 MGICITGATGSGTGTGTVVVARHPKQYVALTANGNDALYECIAHHPYAAVWMS 60
 QY 61 KVAEFKRIIAASPVADIKVLSGSEALQOVATLENVDVTMAAIVGAAGLPTLAAKAGKT 120

DB 61 KVAEFKRIIAASPVADIKVLSGSEALQOVATLENVDVTMAAIVGAAGLPTLAAKAGKT 120
 QY 121 VLANKREALVMSGOIFMQAVSDSGAVLLPIDSEHNAIFQCPAGCTPGHTAKOARRILT 180
 DB 121 VLANKREALVMSGOIFMQAVSDSGAVLLPIDSEHNAIFQCPAGCTPGHTAKOARRILT 180
 QY 181 ASGGPFRTPPIETLSSTVTPDQAVAHPRMDWKRKISVDSATMNGELIEACLPNNRPD 240
 DB 181 ASGGPFRTPPIETLSSTVTPDQAVAHPRMDWKRKISVDSATMNGELIEACLPNNRPD 240
 QY 241 QIEVVIHPQSIITHSMVDYVDSVLAQMGNPDMKRTPIAHAMAPRFDGVAPLDIFEVGH 300
 DB 241 QIEVVIHPQSIITHSMVDYVDSVLAQMGNPDMKRTPIAHAMAPRFDGVAPLDIFEVGH 300
 QY 301 MDEKPDLRPPCLRLAYEALIKSGIMPTVLANANEIAYEALNEBEVKETDIAYIIERSM 360
 DB 301 MDEKPDLRPPCLRLAYEALIKSGIMPTVLANANEIAYEALNEBEVKETDIAYIIERSM 360
 QY 361 AQFKPDDAGSLLEVLQADQDAREVARDIIKTIVA 394
 DB 361 AQFKPDDAGSLLEVLQADQDAREVARDIIKTIVA 394

RESULT 4
 AA021884
 ID AA021884 standard; Protein; 396 AA.
 XX AA021884;
 AC
 XX 13-SEP-2002 (first entry)
 DT
 XX Isoprenoid related protein sequence SEQ ID No 117.
 DE
 XX Isoprenoid related protein sequence SEQ ID No 117.
 KM Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;
 KM decaprenyl diphosphate synthase.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN MO200226933-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 28-SEP-2001; 2001WO-US30328.
 XX
 PR 29-SEP-2000; 2000US-236580P.
 XX
 XX (CRGI) CARGILL INC.
 XX
 PI Gokarn R, Jessen H, Zidwick MJ;
 PI WPI: 2002-416480/44.
 XX
 PT Substantially pure polypeptides having e.5.''
 PT 1-deoxyxylulose-5-phosphate synthase activity, useful for the
 PT production of isoprenoids, especially CoQ(10).
 XX
 PS Disclosure; Fig 32; 246pp; English.
 XX
 CC The invention relates to methods and materials for the production of
 CC isoprenoids. More particularly the invention provides isolated nucleic
 CC acids, substantially pure polypeptides, host cells, and methods for
 CC producing various isoprenoid compounds. The polypeptides are useful for
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)
 CC activity or decaprenyl diphosphate synthase (DDS) activity, is useful for
 CC increasing production of CoQ(10) in a cell having endogenous DDS
 CC activity. This sequence represents a protein relating to the isoprenoid
 CC production of the invention.
 XX
 XX
 SQ Sequence 396 AA;

Query Match 61.2%; Score 1215.5; DB 23; Length 396;

Best Local Similarity 64.1%; Pred. No. 4,1e-110; Matches 245; Conservative 48; Mismatches 88; Indels 1; Gaps 1;

```

QY 4 ICLGATGSGVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVMEKVA 63
DB 7 ISVLGATGSGVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVMEKVA 66
QY 64 EFKGRIASPVDIKVLSGSEALQOVATLENVDYMAAIVGAAGLLPTLAAKAKTIVL 123
DB 67 ALQGSILAAAGIR-TRVLFGEOALCEVASAPEVDMAAIVGAAGLLPTLAAKAKTIVL 125
QY 124 ANKEALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCMPPAGYTPGHAKQARRILLTASG 183
DB 126 ANKEALVMSGALFMQAVKSGAVLLPIDSEHNAIFQSLPRNADGLERGVARRILLTASG 185
QY 184 GPFRRPIETLSVTPDOAVAHPKMDMGRKISVDSATMNNKGLLEIACTLFNNEPDQIE 243
DB 186 GPFRRPIETLSVTPDOAVAHPKMDMGRKISVDSATMNNKGLLEIACTLFNNEPDQIE 245
QY 244 VTIHQSTIHSVVDYVDSVLAQKNDPRTPIAHMAMPERPDSGVAPLDIFEVGMDF 303
DB 246 VTIHQSTIHSVVDYVDSVLAQKNDPRTPIAHMAMPERPDSGVAPLDIFEVGMDF 305
QY 304 EKPDLKFPCLRLAYEAIKSGIMPTVIANANEIAVEAFINEEVEFTDIAVLIERSMAQF 363
DB 306 QKPDQRFPCIKLALAYEAMNAGAPCVIANANEIAVEAFINEEVEFTDIAVLIERSMAQF 365
QY 364 KPDAGSLLEVLAQDQDAREVA 385
DB 366 AVTAVESLDQVLAADRRASVA 387

```

RESULT 5

AAO21886
ID AAO21886 standard; Protein; 394 AA.

AAO21886;

13-SEP-2002 (first entry)

Isoprenoid related protein sequence SEQ ID No 119.

Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;

Neisseria meningitidis.

MO200226933-A2.

04-APR-2002.

28-SEP-2001; 2001WO-US30328.

29-SEP-2000; 2000US-236580P.

(CRGI) CARGILL INC.

Gokarn R, Jessen H, Zidwick MJ;

WPI; 2002-416480/44.

Substantially pure polypeptides having e.g.,

1-deoxyxylulose-5-phosphate synthase activity, useful for the

production of isoprenoids, especially CoQ(10)

production of isoprenoids, especially CoQ(10)

production of isoprenoids, especially CoQ(10)

production of isoprenoids, especially CoQ(10)

production of isoprenoids, especially CoQ(10)

production of isoprenoids, especially CoQ(10)

production of isoprenoids, especially CoQ(10)

activity or decaprenyl diphosphate synthase (DDS) activity, is useful for increasing production of CoQ(10) in a cell having endogenous DDS activity. This sequence represents a protein relating to the isoprenoid production of the invention.

Sequence 394 AA;

Query Match 53.0%; Score 1052.5; DB 23; Length 394;

Best Local Similarity 55.5%; Pred. No. 3.7e-94; Matches 216; Conservative 49; Mismatches 123; Indels 1; Gaps 1;

```

QY 4 ICLGATGSGVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVMEKVA 63
DB 6 LTIAGTSGVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVMEKVA 65
QY 64 EFKGRIASPVDIKVLSGSEALQOVATLENVDYMAAIVGAAGLLPTLAAKAKTIVL 123
DB 66 RLBALIKRDGTA-QVTLHGAQALVDVASADEVSGVMCAIVGAVGLPSALAAQCKTIVL 124
QY 124 ANKEALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCMPPAGYTPGHAKQARRILLTASG 183
DB 125 ANKEALVMSGALFMQAVKSGAVLLPIDSEHNAIFQSLPRNADGLERGVARRILLTASG 184
QY 184 GPFRRPIETLSVTPDOAVAHPKMDMGRKISVDSATMNNKGLLEIACTLFNNEPDQIE 243
DB 185 GPFRRPIETLSVTPDOAVAHPKMDMGRKISVDSATMNNKGLLEIACTLFNNEPDQIE 244
QY 244 VTIHQSTIHSVVDYVDSVLAQKNDPRTPIAHMAMPERPDSGVAPLDIFEVGMDF 303
DB 245 VTIHQSTIHSVVDYVDSVLAQKNDPRTPIAHMAMPERPDSGVAPLDIFEVGMDF 304
QY 304 EKPDLKFPCLRLAYEAIKSGIMPTVIANANEIAVEAFINEEVEFTDIAVLIERSMAQF 363
DB 305 QKPDQRFPCIKLALAYEAMNAGAPCVIANANEIAVEAFINEEVEFTDIAVLIERSMAQF 364
QY 364 KPDAGSLLEVLAQDQDAREVA 392
DB 365 FSDGIDGIGLLAODARTRAQARAFICTL 393

```

RESULT 6

AAO21878
ID AAO21878 standard; Protein; 398 AA.

AAO21878;

13-SEP-2002 (first entry)

Isoprenoid related protein sequence SEQ ID No 100.

Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;

decaprenyl diphosphate synthase.

Escherichia coli.

MO200226933-A2.

04-APR-2002.

28-SEP-2001; 2001WO-US30328.

29-SEP-2000; 2000US-236580P.

(CRGI) CARGILL INC.

Gokarn R, Jessen H, Zidwick MJ;

WPI; 2002-416480/44.

Substantially pure polypeptides having e.g.,

1-deoxyxylulose-5-phosphate synthase activity, useful for the

production of isoprenoids, especially CoQ(10)

production of isoprenoids, especially CoQ(10)

production of isoprenoids, especially CoQ(10)

PS Disclosure; Fig 27; 246pp; English.

CC The invention relates to methods and materials for the production of
 CC isoprenoids. More particularly the invention provides isolated nucleic
 CC acids, substantially pure polypeptides, host cells, and methods for
 CC producing various isoprenoid compounds. The polypeptides are useful for
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (Dxs)
 CC activity or decaprenyl diphosphate synthase (Dds) activity, is useful for
 CC increasing production of CoQ(10) in a cell having endogenous Dds
 CC activity. This sequence represents a protein relating to the isoprenoid
 CC production of the invention.

CC Sequence 398 AA;

Query Match 52.6%; Score 1045; DB 23; Length 398;
 Best Local Similarity 54.9%; Pred. No. 2e-93;
 Matches 219; Conservative 56; Mismatches 118; Indels 6; Gaps 3;

QY 1 MKGICILGATSGISGVTLDVVARHPDKYQVVALTMANGNIDALYEQCLAHPEYAVVMES 60
 DB 1 MKQITLIGSTGICSTLDVVRHNEHFRVVALVAGKVTREVTQCLEFSPRYAVMDDEA 60
 QY 61 KVAERKORIASPVADIKVLSGSEBALQOVATLENDVTMAAIVSAGLPTLAAKAGKT 120
 DB 61 S-AKLKTLVLOQGSRTVELSGQAACTMALEDDVDQVMAAIVSAGLPTLAAIRAGKT 119
 QY 121 VLLANKREALVMSGOIFMOAVSDSGAVLLPIDSEHNAIFQCM--PAGYTPGHTAKQAR-- 175
 DB 120 ILLANKESIVTCGRFMDAVKQSKQALLPVDSEHNAIFQSLPQPIQHNIGVADLEQNGV 179
 QY 176 RILLTASGPPRRPTPIETLSSVTPOQAVAHPRKMDGRIKISVDSATMKNKGLEIEACILF 235
 DB 180 SILTIGSGGPPRETPRLDALTPTDQACRHPNMSGRIKISVDSATMKNKGLEYEARWLF 239
 QY 236 NMEPOIEVVIHQSIHSMVDYVDSVLAQMGNDPMTPIAHAAAMPERPDSGVAAPDI 295
 DB 240 NASAQMEVLIHQSVIHSWRYODGSVLAQIGSPDMETPIAHMAAPNPNRVNSGVKPLDF 299
 QY 296 FEVGHMDPEKPDLPKRPCLRLAYEAIKSGGIMPTVLAANEIAYEAPLNEEYKFTDIAT 355
 DB 300 CKLSALTFAAPDYDRYPCIKLAMEAFEGQAATTAALNAANEITVAAPLAAQOIRFTDIAL 359
 QY 356 IERSMAQFKPDDAGSILEVLAQDODAREVARDIKTLVA 394
 DB 360 NLSVLERKMDKREPQCDVVLSDVNAAREVARKVEYRLAS 398
 RESULT 7
 ID AAO21885 strand; Protein; 398 AA.
 AC AAO21885;
 AC AAO21885;
 DT 13-SEP-2002 (first entry)
 DE Isoprenoid related protein sequence SEQ ID No 118.
 DE Isoprenoid related protein sequence SEQ ID No 118.
 KM Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; Dxs; Dds;
 KM decaprenyl diphosphate synthase.
 OS Escherichia coli.
 OS Escherichia coli.
 XX WO200226933-A2.
 XX WO200226933-A2.
 PD 04-APR-2002.
 XX 28-SEP-2001; 2001WO-US30328.
 XX 29-SEP-2000; 2000US-236580P.
 XX (CRGI) CARGILL INC.
 XX

PI Gokarn R, Jessen H, Zidwick MJ;
 XX 'DR WPI; 2002-416480/44.
 XX

PT Substantially pure polypeptides having e.g.,
 PT 1-deoxyxylulose-5-phosphate synthase activity, useful for the
 PT production of isoprenoids, especially CoQ(10)
 PS Disclosure; Fig 32; 246pp; English.

CC The invention relates to methods and materials for the production of
 CC isoprenoids. More particularly the invention provides isolated nucleic
 CC acids, substantially pure polypeptides, host cells, and methods for
 CC producing various isoprenoid compounds. The polypeptides are useful for
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (Dxs)
 CC activity or decaprenyl diphosphate synthase (Dds) activity, is useful for
 CC increasing production of CoQ(10) in a cell having endogenous Dds
 CC activity. This sequence represents a protein relating to the isoprenoid
 CC production of the invention.

CC Sequence 398 AA;

Query Match 52.6%; Score 1045; DB 23; Length 398;
 Best Local Similarity 54.9%; Pred. No. 2e-93;
 Matches 219; Conservative 56; Mismatches 118; Indels 6; Gaps 3;

QY 1 MKGICILGATSGISGVTLDVVARHPDKYQVVALTMANGNIDALYEQCLAHPEYAVVMES 60
 DB 1 MKQITLIGSTGICSTLDVVRHNEHFRVVALVAGKVTREVTQCLEFSPRYAVMDDEA 60
 QY 61 KVAERKORIASPVADIKVLSGSEBALQOVATLENDVTMAAIVSAGLPTLAAKAGKT 120
 DB 61 S-AKLKTLVLOQGSRTVELSGQAACTMALEDDVDQVMAAIVSAGLPTLAAIRAGKT 119
 QY 121 VLLANKREALVMSGOIFMOAVSDSGAVLLPIDSEHNAIFQCM--PAGYTPGHTAKQAR-- 175
 DB 120 ILLANKESIVTCGRFMDAVKQSKQALLPVDSEHNAIFQSLPQPIQHNIGVADLEQNGV 179
 QY 176 RILLTASGPPRRPTPIETLSSVTPOQAVAHPRKMDGRIKISVDSATMKNKGLEIEACILF 235
 DB 180 SILTIGSGGPPRETPRLDALTPTDQACRHPNMSGRIKISVDSATMKNKGLEYEARWLF 239
 QY 236 NMEPOIEVVIHQSIHSMVDYVDSVLAQMGNDPMTPIAHAAAMPERPDSGVAAPDI 295
 DB 240 NASAQMEVLIHQSVIHSWRYODGSVLAQIGSPDMETPIAHMAAPNPNRVNSGVKPLDF 299
 QY 296 FEVGHMDPEKPDLPKRPCLRLAYEAIKSGGIMPTVLAANEIAYEAPLNEEYKFTDIAT 355
 DB 300 CKLSALTFAAPDYDRYPCIKLAMEAFEGQAATTAALNAANEITVAAPLAAQOIRFTDIAL 359
 QY 356 IERSMAQFKPDDAGSILEVLAQDODAREVARDIKTLVA 394
 DB 360 NLSVLERKMDKREPQCDVVLSDVNAAREVARKVEYRLAS 398
 RESULT 8
 ID ABP79699 strand; Protein; 411 AA.
 AC ABP79699;
 AC ABP79699;
 DT 07-MAR-2003 (first entry)
 DE N. gonorrhoeae amino acid sequence SEQ ID 5928.
 DE N. gonorrhoeae amino acid sequence SEQ ID 5928.
 KM Antibacterial; infection; vaccine; gene therapy.
 KM Antibacterial; infection; vaccine; gene therapy.
 OS Neisseria gonorrhoeae.
 OS Neisseria gonorrhoeae.
 XX WO200279243-A2.
 XX WO200279243-A2.
 PD 10-OCT-2002.
 PD

XX	12-FEB-2002; 2002WO-IB02069.
PF	
XX	12-FEB-2001, 2001GB-0003424.
PR	
XX	(CHIR-) CHIRON SPA.
PA	
XX	
XX	Fontana MR, Piazza M, Masignani V, Monaci B;
PI	
XX	WPI; 2003-058415/05.
DR	
XX	N-PSDE; ABZ40669.
XX	
PT	New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT	medicament for treating or preventing N. gonorrhoeae infection
PS	
XX	Disclosure; Page 621; 815pp; English.
XX	
CC	The present invention relates to proteins from Neisseria gonorrhoeae.
CC	Also disclosed are the nucleic acid molecules encoding the proteins and
CC	antibodies that specifically bind to the proteins. The composition
CC	comprising the protein, nucleic acid or antibody is useful for the
CC	manufacture of a medicament for treating or preventing N. gonorrhoeae
CC	infection, this may be in the form of a vaccine or gene therapy.
CC	Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC	molecules of the invention.
XX	
XX	
Sequence	411 AA;

Query Match	52.6%	Score 1043.5	DB 24,	Length 411,
Best Local Similarity	55.5%	Pred. No. 3e-03		
Matches 216,	Conservative 48,	Mismatches 124,	Indels 1,	Gaps 1,

QY	4	ICILATGTSIGVSTEDVVAHRPDKTQVVALTANGIDLYEQCLAHHPREVAVVMESKVA	63
Db	23	LTTIGSTGSI GEBSTLDVSRHPKRRVYALASHQVETLAAQCOCTPRPRYAVVADAEHNA	82
QY	64	EFFQRIASBPADIKYLSGSSEALQVATLENTVYMAAIVAGAGLLPTLAAARGKTVLL	123
Db	83	RLIALIKPQSTA -TQYHGAQALVVASDEVSQWCAIVGAALPBLALMAAOKRGKTVY	141
QY	124	ANKEALVMSQIIMQAVSDSGAVLLPISSEHNAIFQCPACATYTPGHRKXQARELLTASG	183
Db	142	ANKEETLVSGSALFMETAPRANGAVALPVPSEHNAIFQVLPDYDTLRLEHGISDILLTASG	201
QY	184	GPRFRTPIFLTSVSTPDQAAVAPKMDMGKISVSDSATVANKGELIEACLLFNMPDQTE	243
Db	202	GPELTIDLSFTSDITPEQAVKFPNRMGRKISVDSATVANKGELIEAHNLPNCPPKYLE	261
QY	244	VIIHPQSIHSWVDYVDSVLAQMGNPDMKTRPIAHAMKAMPERPDSVAPLIDFEVGHMDF	303
Db	262	VIIHPQSVIHSWVRKSDGSLAQLGNDPMKTRPIAVCYLLPERIDSGVGKIDFGALSALTPE	321
QY	304	EKEDLKRPFLCLALAYALAKSGGIMPLVYLAAMELVAFLMEVYKFTDIAVIERSSAQC	363
Db	322	QKPDFFGFPCILKFAVETINAGGAAPCVYLAANAEHTVAAFDGOIKFTDIAKTVAHCLAQD	381
QY	364	KPDAGSLLEYLQADQDAREVARDIKTL	392
Db	382	FSNGMDIISGLAQDARTBAQARAFICTL	410

RESULT 9
ABP80486
ID ABP80486 standard; Protein; 411 AA.

AC ABP80486;

DT 07-MAR-2003 (First entry)

DB N. gonorrhoeae amino acid sequence SEQ ID 7502.

KW Antibacterial; infection; vaccine; gene therapy.

OS *Neisseria gonorrhoeae*.

PN WO200279243-A2

PD 10-OCT-2002

PF 12-FEB-2002; 2002WO-1B02069.

PR 12-FEB-2001; 2001GB-0003424.

PA (CHIR-) CHIRON SPA.

PI Fontana MR, Pizza M, Massignani V, Monaci E;

DR WPT: 2003-058415/

XXXX

PT medicament for treating or preventing N. gonorrhoeae infection -

PS Disclosure; Page 734; 815pp; English

The present invention relates to proteins from *Neisseria gonorrhoeae*. Also disclosed are the nucleic acid molecules encoding the proteins and antibodies that specifically bind to the proteins. The composition comprising the protein, nucleic acid or antibody is useful for the manufacture of a medicament for treating or preventing *N. gonorrhoeae* infection, this may be in the form of a vaccine or gene therapy. Sequences given in records ABP6736-ABP8106 represent nucleic acid molecules of the invention.

SQ Sequence 411 AA;

Query Match	52.6%	Score	1043.5	DB	24	Length	411
Best Local Similarity	55.5%	Pred. No.	3e+93				
Matches	16	Conservative	48	Mismatches	12	Indels	1
						Gaps	1

```
QY      4 ICLTATGTSIGSTLDVVAHHDKXOVALLANONIDALTEOCCLAHHEPAAVVMMESKVA    63
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      23 LTLTGSSIGSTLDDVSRRHEKRVALGHGVOKELAAOCOTFFPEPVAVADAEHAH    82

        64 EFKORIAASPVADIKVLSGEALOOVATLENVDVTMAIVAGACILPTLAAAKGTVL    123
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      83 RLRLALKRBDGT-TOLVHGACQLVDVASADESVGMCAIVGAAPSPALAAOCQGIYL    141

        124 ANKEALWMSGOAIFMOAVSDSGAVLLPIDSENAIFOCMPAGYEGTHAXAQRILITASG    183
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      142 ANKETIIVAGGLFMETARANGAAVUPDSENNALFOVLPROYTDRNLMEHGDISTILLASG    204

        184 GPFRRTPILTSSVPDDQAVAHPKMDGRKISVDSATMANKGLELIBACTLLFNMNEPOIB    243
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      202 GPFLTTDLSTFPDSIPPEQAVKHPPNMRMRKISVSSATMANNGLELIBAHMLFNCPBPKLE    261

        244 VVTHPOSTIHSWVDYDGSVLAOMGNPMKRTPIIAHAMAMPERSFGAALPDIEFGHMDF    303
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      262 VVTHPSVSHSWVRKRDSDVLAQCNRPDKRTPIAYCGLPERIDSGVKULDFGALSALTFF    321

        304 EKFDLRPFCLRLAYEAIKSGINPTVTAANEIAVEAFLEAVEKYFTDIAVIERSMAOF    365
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      322 QXDFRFRPCLRKFAVTTIAGAACAACVTANAETAIVAALFDGQIKTFDIALATAVHCIAOD    381
```

Qy 364 KPDDAGSLVLQADQDAREVADIITKL 392
: | : | | | |
Db 382 FSNMGSDIEGLLAQDARTRAQAFAIGTL 410

RESULT 10

ID AAO21889 standard; Protein; 400 AA.

AC AA021889;

DT 13-SEP-2002 (first entry)

XX Isoprenoid related protein sequence SEQ ID No 122.
 DE Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; Dds;
 XX decaprenyl diphosphate synthase.
 KW Pasteurella multocida.
 XX OS
 XX WO200226933-A2.
 XX PD
 XX 04-APR-2002.
 XX PF
 XX 28-SEP-2001; 2001WO-US30328.
 XX PR
 XX 29-SEP-2000; 2000US-236580P.
 XX (CRGI) CARGILL INC.
 XX Gokarn R, Jeesen H, Zidwick MJ;
 XX WPI; 2002-416480/44.
 XX Substantially pure polypeptides having e.g.,
 PT 1-deoxyxylulose-5-phosphate synthase activity, useful for the
 PT production of isoprenoids, especially CoQ(10).
 XX Disclosure; Fig 32; 246pp; English.
 XX The invention relates to methods and materials for the production of
 CC isoprenoids. More particularly the invention provides isolated nucleic
 CC acids, substantially pure polypeptides, host cells, and methods for
 CC producing various isoprenoid compounds. The polypeptides are useful for
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)
 CC activity or decaprenyl diphosphate synthase (Dds) activity, is useful for
 CC increasing production of CoQ(10) in a cell having endogenous Dds
 CC activity. This sequence represents a protein relating to the isoprenoid
 CC production of the invention.
 XX SQ Sequence 400 AA;
 XX
 XX Query Match 52.5%; Score 1042; DB 23; Length 400;
 XX Best Local Similarity 53.2%; Pred. No. 4,1e-93;
 XX Matches 210; Conservative 68; Mismatches 109; Indels 8; Gaps 3;
 QY 1 MGICITAGTSGTIGTLDVVARHPDKQVVALTANGNIDALYECGLAHHPYAVVWES 60
 DB 2 MKKEVILSTGSIIGTSTLVITHNPDKIVFALVGGRNVELMFQCLTFQPSFALDDV 61
 QY 61 KVAEFKORIAAPVADIKVLSGSEALQVATLENVDTVAALVGAAGLLPTLAARAGKT 120
 DB 62 AAKMLAKRKAAHQ-SGTVLAGQQAICELAAHPADMVAALVGAAGLLPTLSAVKAGR 120
 QY 121 VILANKKALVMSGQIFMVAVSDSGAVLPIIDSEHNAIFQCM-----AGTTPGHTAKQA 174
 DB 121 VILANKKALVLTGQGFIDAVRESQAQLPVDSEHNAIFQSLPPEAROGTFCP-ISELGI 179
 QY 175 RRIILTAGSGPRRTPIETLSVTPOAVAAHPKMDGRKISVDSATMKNKGLIEACILF 234
 DB 180 SKIVLTGSGPRRTPIETLSVTPOAVAAHPKMDGRKISVDSATMKNKGLIEARML 239
 QY 235 FNNEPDQIEVVIHPQSIHSWVDYDGSYLAQMGNDKMTPIAHAMAFERDSCVAPLD 294
 DB 240 FNASAEEMEVIIHPQSIHSWVDYDGSYLAQMGNDKMTPIAHAMAFERDSCVAPLD 299
 QY 295 IEVHGNDKMDKRPCLRLAYEAIKSGIMPTLVANANIEAVEAFINBEVKFTDIAY 354
 DB 300 FQJNLNLTIEPDYQYRPLKLAIDAFSGVARTTMAANAIEIAVASFIDMKITETDIAR 359
 QY 355 IIERNSAOPKPDAGSLLEVLQAQDAREVARDII 389
 DB 360 LMQLVVSKLOPKIHCIEDVLAVDKKARELSQII 394

RESULT 11
 ID AAO21877
 AC AAO21877 standard; Protein; 397 AA.
 XX
 XX AAO21877;
 XX DT
 XX 13-SEP-2002 (first entry)
 XX DE
 XX Isoprenoid related protein sequence SEQ ID No 99.
 XX Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; Dds;
 KW decaprenyl diphosphate synthase.
 XX OS
 XX Haemophilus influenza.
 XX WO200226933-A2.
 XX PN
 XX 04-APR-2002.
 XX PF
 XX 28-SEP-2001; 2001WO-US30328.
 XX PR
 XX 29-SEP-2000; 2000US-236580P.
 XX (CRGI) CARGILL INC.
 XX Gokarn R, Jeesen H, Zidwick MJ;
 XX WPI; 2002-416480/44.
 XX Substantially pure polypeptides having e.g.,
 PT 1-deoxyxylulose-5-phosphate synthase activity, useful for the
 PT production of isoprenoids, especially CoQ(10).
 XX Disclosure; Fig 27; 246pp; English.
 XX The invention relates to methods and materials for the production of
 CC isoprenoids. More particularly the invention provides isolated nucleic
 CC acids, substantially pure polypeptides, host cells, and methods for
 CC producing various isoprenoid compounds. The polypeptides are useful for
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)
 CC activity or decaprenyl diphosphate synthase (Dds) activity, is useful for
 CC increasing production of CoQ(10) in a cell having endogenous Dds
 CC activity. This sequence represents a protein relating to the isoprenoid
 CC production of the invention.
 XX SQ Sequence 397 AA;
 XX
 XX Query Match 50.7%; Score 1007; DB 23; Length 397;
 XX Best Local Similarity 50.6%; Pred. No. 1,1e-89;
 XX Matches 200; Conservative 76; Mismatches 111; Indels 8; Gaps 3;
 QY 2 KGICITAGTSGTIGTLDVVARHPDKQVVALTANGNIDALYECGLAHHPYAVVWESX 61
 DB 4 QNIVILSTGSIIGTSTLVITHNPDKIVFALVGGRNVELMFQCLTFQPSFALDDVNA 63
 QY 62 VAEKFORIAAPVADIKVLSGSEALQVATLENVDTVAALVGAAGLLPTLAARAGKT 121
 DB 64 AKIIRKLIHNI-PTETLAGRAICELAAHPADMVAALVGAAGLLPTLSAVKAGR 122
 QY 122 LANKKALVMSGQIFMVAVSDSGAVLPIIDSEHNAIFQCM-----AGTTPGHTAKQA 175
 DB 123 LANKKALVLTGQGFIDAVKNGYGLLPVDSEHNAIFQSLPPEAROGTFCP-ISELGV 181
 QY 176 RRIILTAGSGPRRTPIETLSVTPOAVAAHPKMDGRKISVDSATMKNKGLIEACILF 235
 DB 182 KIILTSGSGPRRTPIETLSVTPOAVAAHPKMDGRKISVDSATMKNKGLIEARML 241
 QY 236 FNNEPDQIEVVIHPQSIHSWVDYDGSYLAQMGNDKMTPIAHAMAFERDSCVAPLDI 295
 DB 242 NASAEEMEVIIHPQSIHSWVDYDGSYLAQMGNDKMTPIAHAMAFERDSCVAPLDI 301

QY 296 FEVGHMDPEKDPKRPCLRLAYEAIRKSGIMPTVLNANEIAYEAPLINEVEKTDIAVI 355
 DB 302 FKIKELTFIEPDPENRYPNLKLAIDAFAGAYATTAMNANEIAYQAFIDRQIGFMDIAKI 361
 QY 356 IERSMAQFEPDAGSLLEVLQADQDAREVARDIIR 390
 DB 362 NSKTERISPTTIOINIDVLEIDAQAREIAKTILIR 396

RESULT 12

AAO21887
 ID AAO21887 standard; Protein; 397 AA.

AAO21887;

13-SEP-2002 (first entry)

Isoprenoid related protein sequence SEQ ID No 120.

Isoprenoid, CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;
 decaprenyl diphosphate synthase.

Haemophilus influenza.

MO200226933-A2.

04-APR-2002.

28-SEP-2001; 2001MO-US30328.

29-SEP-2000; 2000US-236580P.

(CRGI) CARGILL INC.

Gokarn R, Jessen H, Zidwick MJ;

WPI; 2002-416480/44.

Substantially pure polypeptides having e.g.,
 1-deoxyxylulose-5-phosphate synthase activity, useful for the
 production of isoprenoids, especially CoQ(10)

Disclosure: Fig 32, 246pp; English.

The invention relates to methods and materials for the production of
 isoprenoids. More particularly the invention provides isolated nucleic
 acids, substantially pure polypeptides, host cells, and methods for
 producing various isoprenoid compounds. The polypeptides are useful for
 the production of isoprenoids, especially CoQ(10). Expressing the pure
 polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)
 activity or decaprenyl diphosphate synthase (DDS) activity, is useful for
 increasing production of CoQ(10) in a cell having endogenous DDS
 activity. This sequence represents a protein relating to the isoprenoid
 production of the invention.

Sequence 397 AA;

Query Match 50.7%; Score 1007; DB 23; Length 397;

Best Local Similarity 50.6%; Pred. No. 1,1e-89;

Matches 200; Conservative 76; Mismatches 111; Indels 8; Gaps 3;

2 KGICILCATSGISGVTLIDVVARHPDKQVVALNKNIDILYQCLAHNEHYAVVMESEK 61

4 QNIVILGISTSIGSTISVLENNPQKHAFALVGGKVMZMFQCKIFRHFALDDVNA 63

62 VAEFKORIASPVADIVLGSSEALQOVATLENDVMAIVAGLPTLAATAKAKTY 121

64 AKILREGLIAHT-PTREVLAGRACILAHAPDADQIMATIVAGLPTLSAVKSKRY 122

122 LLANKELVMSGQIFQAVSDSGAVILPITSEENATFQCM-----AGTTPEHTAKQAR 175

123 LLANKESLTVCGQFLIDAVKVKSGKLLPVDSEHNAIFQSLPFAQEKIGFCP-LSEIGVS 181

QY 176 RILITNAGGPPRRPTPIITLSTVPDQAVAEKMDMGKXISVDSATNMKGLIELIACLP 235
 DB 182 KILITGGGGPRRYPLPFOPTNITPEQAVAHENMGKKISVDSATNMKGLIYEAWLP 241
 QY 236 NMEPDQIEVVIHPOSIHSMVDYDGSVLQAGMPDKRTPIAHAMAPREPSGVAPLDI 295
 DB 242 NASAEHEVVIHPOSIHSMRYDGSVITQMGNDPKRTPIAHAMAPREPSGVAPLDI 301
 QY 296 FEVGHMDPEKDPKRPCLRLAYEAIRKSGIMPTVLNANEIAYEAPLINEVEKTDIAVI 355
 DB 302 FKIKELTFIEPDPENRYPNLKLAIDAFAGAYATTAMNANEIAYQAFIDRQIGFMDIAKI 361
 QY 356 IERSMAQFEPDAGSLLEVLQADQDAREVARDIIR 390
 DB 362 NSKTERISPTTIOINIDVLEIDAQAREIAKTILIR 396

RESULT 13

ABB09436
 ID ABB09436 standard; Protein; 397 AA.

ABB09436;

01-JUL-2002 (first entry)

H. influenzae DXR reductoisomerase enzyme polypeptide sequence.

DXR, reductoisomerase; enzyme; non-mevalonate isoprenoid;

meningitis; pneumonia; conjunctivitis; bacteremia; sinusitis;

pleural empyema; endocarditis; epiglottitis.

Haemophilus influenzae.

MO200211673-A2.

14-FEB-2002.

09-AUG-2001; 2001MO-US24950.

09-AUG-2000; 2000US-223909P.

(SMIR) SMITHKLINE BEECHAM CORP.

(SMIR) SMITHKLINE BEECHAM PLC.

Jaworski DD, Payne DJ, Slater-Radostci CE, Yan K;

WPI; 2002-241698/29.

Modulating Haemophilus influenzae DXR reductoisomerase enzyme activity,
 useful for treating mammals or tissues infected with H. influenzae
 (e.g. ear infections or pneumonia) by contacting the enzyme with a
 modulator of its activity

Claim 6(i); Page 5; 44pp; English.

The invention relates to modulating an activity of a DXR reductoisomerase
 enzyme of Haemophilus influenzae, comprising contacting the enzyme with a
 compound that modulates non-mevalonate isoprenoid biosynthesis -
 synthesis of menaquinone or ubiquinone. Compounds of the invention act as
 vinocides. The method is useful for treating a mammal or mammalian tissue
 infected with H. influenzae having DXR reductoisomerase enzyme, e.g. a
 human or a domestic animal. In particular, the method is useful for
 treating ear infections, conjunctivitis, meningitis, pneumonia,
 conjunctivitis, bacteremia, sinusitis, pleural empyema, endocarditis and
 epiglottitis. The current sequence represents the H. influenzae DXR
 reductoisomerase enzyme polypeptide sequence.

Note: in the sequence listing, on page 40-44 of the specification, there
 is given an amino acid sequence that is also described as SEQ ID 2, which
 is the identifier of the current sequence. This sequence is given in
 record ABB09437, and contains the amino acids given in the current
 sequence but with a large insertion of extra amino acids that are not

CC encoded in the polynucleotide that encodes the enzyme of the invention
 CC (see AB40400).

XX Sequence 397 AA;

Query Match 50.7%; Score 1007; DB 23; Length 397;

Best Local Similarity 50.6%; Pred. No. 1.1e-89; Indels 8; Gaps 3;

Matches 200; Conservative 76; Mismatches 111; Indels 8; Gaps 3;

2 KQICIGATGSGIGSTLDVYARHPDKYQVVALTANGNIDALYEQCLAHREYAVVWMSK 61

4 QNIVILGSGISIGKSTLSTYENNPQYTHAFALVGKNNVAMFEQCKRPHRALDDVNA 63

62 VAEFKRIASPVADIKVTSGESEALQOVATLENVDTMAAIVGAAGLPTLAAAKGTV 121

64 AKILREKLAHNI-PREVLAGRAICELAHDPADIMASIVGAAGLPTLSAVVAGKRV 122

122 LLAKEALVMSGQIFQAVSDSGAVLLPTDSEHNAIFQCP-----AGYTPGHTAKQAR 175

123 LLAKEALVTCGQLFDAYKNGSKLLPVDSEHNAIFQSLPPEAQKIGFCP-LSHIGVS 181

176 RILFNASGSPFRPTPIETLSVTPDQVAHAPKMDGKRKISVDSATMNGLELIEACLF 235

182 KIILTGSGSPFRPTPIETLSVTPDQVAHAPKMDGKRKISVDSATMNGLELIEACLF 241

236 NMEPDQIEVVIHPOSIIHSMNDVYDSVLAQKGNPDKRTPIAHAMAPRPSGVAPLDI 295

242 NNAEEMEVIIHPOSIIHSMNDVYDSVLAQKGNPDKRTPIAHAMAPRPSGVAPLDI 301

296 FEVGHNDPEKPKDLKRPCLRLAVEAIIKSGIMPTVLAANEIAVEAFINBEVFTDIATV 355

302 PIKEKELTIEPDRNRPNLKIADPAQVATTANNAEIAVAFLDRQIGFDIATV 361

356 IERSMAQFEPDDAGSELVLAQDQAREVARDIITKVA 390

362 NSKTERISPTIIONIDVLEIDAQREIAKTILR 396

RESULT 14

AAO21879 standard; Protein; 388 AA.

AAO21879;

13-SEP-2002 (first entry)

Isoprenoid related protein sequence SEQ ID No 101.

Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;

decaprenyl diphosphate synthase.

Zymonas mobilis.

MO200226933-A2.

04-APR-2002.

28-SEP-2001; 2001MO-US30328.

29-SEP-2000; 2000US-236580P.

(CRGI) CARGILL INC.

Gokarn R, Jessen H, Zidwick MJ;

WPI; 2002-416480/44.

Substantially pure polypeptides having e.g.,

1-deoxyxylulose-5-phosphate synthase activity, useful for the

production of isoprenoids, especially CoQ(10)

Disclosure; Fig 27; 246pp; English.

CC The invention relates to methods and materials for the production of
 CC isoprenoids. More particularly the invention provides isolated nucleic
 CC acids, substantially pure polypeptides, host cells, and methods for
 CC producing various isoprenoid compounds. The polypeptides are useful for
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)
 CC activity or decaprenyl diphosphate synthase (DDS) activity, is useful for
 CC increasing production of CoQ(10) in a cell having endogenous DDS
 CC activity. This sequence represents a protein relating to the isoprenoid
 CC production of the invention.

XX Sequence 388 AA;

Query Match 50.4%; Score 999.5; DB 23; Length 388;

Best Local Similarity 50.9%; Pred. No. 5.6e-89; Indels 9; Gaps 3;

Matches 199; Conservative 74; Mismatches 109; Indels 9; Gaps 3;

4 ICIGATGSGIGSTLDVYARHPDKYQVVALTANGNIDALYEQCLAHREYAVVWMSKVA 63

7 VYVIGATGSGIGSTLDVYARHPDKYQVVALTANGNIDALYEQCLAHREYAVVWMSKVA 66

64 EFKORIASPVADIKVTSGESEALQOVATLENVDTMAAIVGAAGLPTLAAAKGTV 123

67 DLKEALAS---SYEAAGADALVEAAMW-CADWTMAIIGCAGIKATLAIIRKGTVAL 122

124 ANKEALVMSGQIFQAVSDSGAVLLPTDSEHNAIFQCPAPYTGHTAKQARRIITLTASG 183

123 ANKEALVMSGQIFQAVSDSGAVLLPTDSEHNAIFQCP-----PHNNDYVARIITLTASG 177

184 GPFRTPIETLSVTPDQVAHAPKMDGKRKISVDSATMNGLELIEACLFENNEPDQIE 243

178 GPFRTPIETLSVTPDQVAHAPKMDGKRKISVDSATMNGLELIEACLFENNEPDQIE 237

244 VVIHPOSIIHSMNDVYDSVLAQKGNPDKRTPIAHAMAPRPSGVAPLDIPEVGHNDP 303

238 ILVHPOSIIHSMNDVYDSVLAQKGNPDKRTPIAHAMAPRPSGVAPLDIPEVGHNDP 297

304 EKPDLKRPCLRLAVEAIIKSGIMPTVLAANEIAVEAFINBEVFTDIATVIERMAQF 363

298 EAPDYERFPALTLMESIKSGARPAVANAANETAVAFDLKIKIGFDIATVETKIDHY 357

364 KPDAGSELVLAQDQAREVARDIITKVA 394

358 TPATPSSLEDFALDNEARIQAAALMESLPA 388

RESULT 15

AAO21883 standard; Protein; 388 AA.

AAO21883;

13-SEP-2002 (first entry)

Isoprenoid related protein sequence SEQ ID No 116.

Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;

decaprenyl diphosphate synthase.

Zymonas mobilis.

MO200226933-A2.

04-APR-2002.

28-SEP-2001; 2001MO-US30328.

29-SEP-2000; 2000US-236580P.

(CRGI) CARGILL INC.

Gokarn R, Jessen H, Zidwick MJ;

DR WPI; 2002-416480/44.

XX Substantially pure polypeptides having e.g.,
PT 1-deoxyxylulose-5-phosphate synthase activity, useful for the
PT production of isoprenoids, especially CoQ(10).
XX

PS Disclosure; Fig 32; 246pp; English.

XX
CC The invention relates to methods and materials for the production of
CC isoprenoids. More particularly the invention provides isolated nucleic
CC acids, substantially pure polypeptides, host cells, and methods for
CC producing various isoprenoid compounds. The polypeptides are useful for
CC the production of isoprenoids, especially CoQ(10). Expressing the pure
CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)
CC activity or decaprenyl diphosphate synthase (DPS) activity, is useful for
CC increasing production of CoQ(10) in a cell having endogenous DXS
CC activity. This sequence represents a protein relating to the isoprenoid
CC production of the invention.

XX
SQ Sequence 388 AA;

Query Match 50.4%; Score 999.5; DB 23; Length 388;

Best Local Similarity 50.9%; Pred. No. 5,6e-89;

Matches 199; Conservative 74; Mismatches 109; Indels 9; Gaps 3;

QY 4 ICILGATGSIQVSTLDVVARHPPDKYQVALTANGNIDALYEQCLAHPEYAVVMSKVA 63
DB 7 VVTLGATGSIQVSTLDVVARHPPDKYQVALTANGNIDALYEQCLAHPEYAVVMSKVA 66
QY 64 EFQRILAAFPVADIKVLSGSEALQOVATLENVDVMAAIVGAAGLPTLAAKAGTVLL 123
DB 67 DIKEALAGS---SVEAAGADALVEAAMM-GADMTMAALIGCAGLKATLAAIRKGTVAL 122
QY 124 ANKEALVMSGOLFMOAVSDSGAVLLPIDSEHNAIFQCKPAGYTPGHTAKQARRILLTASG 183
DB 123 ANKESLVSKGLMIDVAREHGTLLPVDEHNAIFQCF----PHARDYVRRIITIASG 177
QY 184 GPFRTPIETLSSVTPDQAVAAHPKMDMGKIKISVDSATWNNKGLLEIACCLFNMEDQIE 243
DB 178 GPFRTTSLAEMATVPERAVOHPNMGAKISIDSATWNNKGLLEIACCLFNMEDQIE 237
QY 244 VVIHQSIHSHWVDVYDGSVLQMGNDPDRTPIAHAMAPERPDGVAPLDIFEVGHDF 303
DB 238 IIVHPOSVTHSWVEYDGSILAQISPDRTPIGHTLAMPKMETPAESLDTKLRQMDP 297
QY 304 EKPDLRPFCLRIAYEAIKSGGIMPTVLNANEIAVEAFINEEVKFTDIIVILERSMAQF 363
DB 298 EAPDYERFPALTLMESIKSGGARPAVMAANEIAVPAFLDKKIGFLDIKIVEXTLDHY 357
QY 364 KPDDAGSLFLYLQADODAREVARDIKTLVA 394
DB 358 TPATPSLSDVFAIDNEARIQAALMESLPA 388

Search completed: January 29, 2004, 15:49:39
Job time : 37.0149 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: January 29, 2004, 15:45:34 (Search time 12.3656 Seconds
(without alignments)
1348.130 Million cell updates/sec

Title: US-09-941-947a-8

Perfect score: 1985
Sequence: 1 MKGICILGATSGIVSTLDV.....IQADQDAREVARDIKTLVA 394

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1215.5	61.2	617	4	US-09-252-991A-21113 Sequence 21113, A
2	1114.5	56.1	398	4	US-09-328-352-7199 Sequence 7199, A
3	818.5	41.2	399	3	US-09-491-362-7 Sequence 7, Appli
4	818.5	41.2	399	4	US-09-874-562-7 Sequence 7, Appli
5	818.5	41.2	477	4	US-09-449-335-6 Sequence 6, Appli
6	817.5	41.2	477	4	US-09-480-921B-10 Sequence 10, Appli
7	814.5	41.0	477	4	US-09-449-335-2 Sequence 2, Appli
8	813.5	41.0	477	4	US-09-480-921B-29 Sequence 29, Appli
9	741	37.3	475	3	US-09-874-562-2 Sequence 2, Appli
10	741	37.3	475	4	US-09-874-562-2 Sequence 2, Appli
11	620	31.2	340	4	US-09-198-452A-362 Sequence 362, App
12	111	5.6	671	4	US-09-252-991A-1862 Sequence 31862, A
13	107	5.4	2285	4	US-09-252-991A-17790 Sequence 17790, A
14	105.5	5.3	3816	4	US-09-428-517-3 Sequence 20932, A
15	99	5.0	3816	3	US-09-428-517-3 Sequence 20932, A
16	99	5.0	390	4	US-09-252-991A-20932 Sequence 17607, A
17	98.5	5.0	394	4	US-09-328-352-4916 Sequence 4916, Ap
18	97	4.9	292	4	US-09-501-115-28 Sequence 28, Appli
19	96.5	4.9	805	4	US-09-252-991A-32752 Sequence 32752, A
20	95	4.8	423	4	US-09-328-352-5657 Sequence 5657, Ap
21	95	4.8	1049	4	US-09-252-991A-25336 Sequence 25336, A
22	94.5	4.8	319	4	US-09-252-991A-50187 Sequence 30187, A
23	94.5	4.8	425	4	US-09-134-001C-4433 Sequence 4433, Ap
24	94	4.7	461	3	US-09-355-115-7 Sequence 7, Appli
25	94	4.7	568	4	US-09-252-991A-22727 Sequence 22727, A
26	93.5	4.7	581	4	US-09-252-991A-25556 Sequence 25556, A
27	93.5	4.7	628	3	US-08-776-271-2 Sequence 2, Appli

28	93.5	4.7	628	3	US-09-215-035-2 Sequence 2, Appli
29	93	4.7	1346	3	US-09-105-537-37 Sequence 37, Appli
30	93	4.7	11877	3	US-09-105-537-6 Sequence 6, Appli
31	92.5	4.7	1562	3	US-09-320-878-3 Sequence 3, Appli
32	92.5	4.7	1562	3	US-09-105-537-35 Sequence 35, Appli
33	92.5	4.7	1562	4	US-09-141-908-4 Sequence 4, Appli
34	92.5	4.7	1562	4	US-09-657-440-3 Sequence 3, Appli
35	91.5	4.6	584	1	US-08-426-819A-36 Sequence 36, Appli
36	91.5	4.6	622	1	US-08-426-819A-35 Sequence 35, Appli
37	91.5	4.6	673	4	US-09-328-352-5834 Sequence 5834, Ap
38	91.5	4.6	710	4	US-09-252-991A-32789 Sequence 32789, A
39	91	4.6	1346	3	US-09-320-878-4 Sequence 4, Appli
40	91	4.6	1346	4	US-09-141-908-5 Sequence 5, Appli
41	91	4.6	1346	4	US-09-657-440-4 Sequence 4, Appli
42	90.5	4.6	752	4	US-09-252-991A-29967 Sequence 29967, A
43	90	4.5	695	6	Patent No. 5460961
44	89.5	4.5	706	4	US-09-252-991A-25730 Sequence 25730, A
45	89.5	4.5	6396	4	US-09-410-551B-72 Sequence 72, Appli

ALIGNMENTS

RESULT 1
US-09-252-991A-21113
Sequence 21113, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
PRT REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21113
LENGTH: 617
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21113

Query Match	Best Local Similarity	Score	DB 4	Length	617
Matches 246	Conservative 47	Mismatches 88	Indels 1	Gaps 1	
4	ICIGATSGIGSTLDVVAHHPDKYQVVALTANGNIDALVEQCLAHHPYAVVMESKVA	63			
228	ISVIGATSGIGSTLDVVAHHPDKYQVVALTANGNIDALVEQCLAHHPYAVVMESKVA	287			
64	EFKQRIASPVADIKYSGSEALQVATLENDVTMAAIYGAALLPTLAAKAGKTYLL	123			
288	ALQGLAAAGR-TRVIFGEQALCEVASAEVDMVAIYGAALLPTLAAKAGKTYLL	346			
124	ANKALVMSQIFMCAVSDGAVLLPDSERNAIFQCPAGYTGHTAKOARRLLTASG	183			
347	ANKALVMSQIFMCAVSDGAVLLPDSERNAIFQCPAGYTGHTAKOARRLLTASG	406			
184	GPFRPTLETLSSVTPDQAVAHPRMDGRKI SVDSATVANKGLEILACLIENKPDIE	243			
407	GPFRPTLETLSSVTPDQAVAHPRMDGRKI SVDSATVANKGLEILACLIENKPDIE	466			
244	VVHPOSIIISMVDYVGVSLAOWKPNDRPTIAHAAWPERPDGVAPLDIFVGHDF	303			
467	VVHPOSIIISMVDYVGVSLAOWKPNDRPTIAHAAWPERPDGVAPLDIFVGHDF	526			
304	EXPDLPKPCRLRLAYEAIKSGGIMPTVLNANELAYEAFINEEYKTDIAVILERSAQF	363			
527	GRPDQGFPCRLRLAYEAIKSGGIMPTVLNANELAYEAFINEEYKTDIAVILERSAQF	586			

QY 364 KPDDAGSLLEVLQADQAREVA 385
 DB 587 AVTAVESLDQVLAADRARSVA 608

RESULT 2

US-09-328-352-7199
 ; Sequence 7199, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 7199
 ; LENGTH: 398
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 ; US-09-328-352-7199

Query Match 56.1%; Score 1114.5; DB 4; Length 398;
 Best Local Similarity 56.7%; Pred. No. 3.4e-114;
 Matches 224; Conservative 58; Mismatches 108; Indels 5; Gaps 2;

QY 2 KGICILGATSGIGSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVNESK 61
 DB 3 QSVCLIGVTVGSIGRSTLTKLIGQHPDKYQVVAHSRISSEIVEIKOPRPK-VVVVPEOK 61
 QY 62 VAEFKRIASPVADIKVLSGSEALQOVATLENDVTMAIYGAAGLPTLAAAKAGKTIV 121
 DB 62 IAEIKTLTPAQQNSIDIDIVLAGGGLVDIASHTDVIWMAIYGAAGLPTLAAVAGARV 121
 QY 122 LIANKEALVMSGOIFMQAVDSGAVLPIDSEHNAIPQCPAGYTPGHTAKQAR---RI 177
 DB 122 LIANKEALVMSGEIMMQARHQAHLPLPDSHNAIPQSLPHNYLQADRTGQPOLGVSKI 181
 QY 178 LITAGGPRRTPIETLSVTPDOAVAHKMDGKISVDSATMNNKGLLEIEACLTNNM 237
 DB 182 LITAGGPRPLNLSHQLTHVTYQACKHNMWNGOGLSDSATLNNKGLLEIEACHLPI 241
 QY 238 BPDQEVVTHPQSIHSMVDYVDSVLAQMGNDPDKRTPIAHAAWPERFDSGVAPLDFE 297
 DB 242 SEHFTVAVHPQSIHSMVQYVDGSLAQMGNDPDKRTPIAHAAWPERLQTVVPALDFE 301
 QY 298 VEHDPKPKDLKRPCLRLAYEAIKSGGIMPTVLANAEIYEAFLNEAVKFTDIAYIE 357
 DB 302 YSQLNFQADDTQKFPALNLAHQMRAGGLAPFILNAANEIYEAFLMERIGETSIPOVE 361
 QY 358 RMAQFKPDDAGSLLEVLQADQAREVARDIKTL 392
 DB 362 HTLEKLENNAAASIEICILDKQVAAVAGQYIISI 396

RESULT 3

US-09-491-362-7
 ; Sequence 7, Application US/09491362
 ; Patent No. 6281017
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney B
 ; APPLICANT: Lange, Bernd M
 ; TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
 ; FILE REFERENCE: WSIU14977
 ; CURRENT APPLICATION NUMBER: US/09/491,362
 ; CURRENT FILING DATE: 2000-01-26
 ; EARLIER APPLICATION NUMBER: 60/118,349
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 7

LENGTH: 399
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-491-362-7

Query Match 41.2%; Score 818.5; DB 3; Length 399;
 Best Local Similarity 44.6%; Pred. No. 1.6e-81;
 Matches 176; Conservative 71; Mismatches 131; Indels 17; Gaps 5;

QY 2 KGICILGATSGIGSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVNESK 61
 DB 3 KPISIVSTSGISGHTLDIVAEKDPKRYVAAAGSVTLADQVRRFKALVAVERESL 62
 QY 62 VAEFKRIASPVADIKVLSGSEALQOVATLENDVTMAIYGAAGLPTLAAAKAGKTIV 121
 DB 63 INELKRALADLDY-KLEIIPGEGVIEVAHPPEAVTVVTVIGGAGKPTVVAIEAGKDI 121
 QY 122 LIANKEALVMSGOIFMQAVDSGAVLPIDSEHNAIPQCPAGYTPGHTAKQARRIL 178
 DB 122 ALANKETLIAGGPRVVLPLANKHNVKILPADSEHSAIPQCIQGLPBG-----ALKIL 173
 QY 179 LITAGGPRRTPIETLSVTPDOAVAHKMDGKISVDSATMNNKGLLEIEACLTNNM 238
 DB 174 LITAGGAFRPMPEKLEKVVADALKHPNMWNGKLTVDSATLNNKGLLEIEATYLGAE 233
 QY 239 PDIIEVTHPQSIHSMVDYVDSVLAQMGNDPDKRTPIAHAAWPERFDSGVAP---LDI 295
 DB 234 YDIEIYTHPQSIHSMVTDSSVLAQMGNDPDKRTPIAHAAWPERFDSGVAP---LDI 293
 QY 296 FEVHMDPEKPKDLKRPCLRLAYEAIKSGGIMPTVLANAEIYEAFLNEAVKFTDIAYI 355
 DB 294 CKIGSLFFKPKPDNKYQVSMOLAYAGAGGTMTGVSAAAEKAVEMFIDEXISYLDIFKV 353
 QY 356 IERSMAQFKPD--DAGSLLEVLQADQAREVARDI 388
 DB 354 VELTCDRHNELVTSPLSEIYVHDTLMAREYAAV 388

RESULT 4

US-09-874-562-7
 ; Sequence 7, Application US/09874562
 ; Patent No. 6420159
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney B
 ; APPLICANT: Lange, Bernd M
 ; TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
 ; FILE REFERENCE: WSIU17549
 ; CURRENT APPLICATION NUMBER: US/09/874,562
 ; CURRENT FILING DATE: 2001-06-04
 ; PRIOR APPLICATION NUMBER: 09/491,362
 ; PRIOR FILING DATE: 2000-01-26
 ; PRIOR APPLICATION NUMBER: 60/118,349
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 7
 ; LENGTH: 399
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-874-562-7

Query Match 41.2%; Score 818.5; DB 4; Length 399;
 Best Local Similarity 44.6%; Pred. No. 1.6e-81;
 Matches 176; Conservative 71; Mismatches 131; Indels 17; Gaps 5;

QY 2 KGICILGATSGIGSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVNESK 61
 DB 3 KPISIVSTSGISGHTLDIVAEKDPKRYVAAAGSVTLADQVRRFKALVAVERESL 62
 QY 62 VAEFKRIASPVADIKVLSGSEALQOVATLENDVTMAIYGAAGLPTLAAAKAGKTIV 121
 DB 63 INELKRALADLDY-KLEIIPGEGVIEVAHPPEAVTVVTVIGGAGKPTVVAIEAGKDI 121

122 LLANKEALVMSGQIFMGAVSDGAVLLPIDSEHNAIFOC---MPAGYTPGHTAKQARRIL 178
122 ALANKETLIAGSPFLPLANKKNVILIPADSEHNAIFOCIOGLPBG-----ALRKII 173
179 LTASGAPRRRTIEFLSSVTPDOAVAHPRKMDGRIKISVDSATMNGKIELLEIACILFPMME 238
174 LTASGAPRRRTIEFLSSVTPDOAVAHPRKMDGRIKISVDSATMNGKIELLEIACILFPMME 233
239 PPOIEVVIHPOSIHSMVDYVDSVLAQMGNDPMTPIAHAMAPRRPDSGVAP---LDI 295
234 YDDIEIVIHPOSIHSMVDYVDSVLAQMGNDPMTPIAHAMAPRRPDSGVAP---LDI 293
236 FEVGNDEPKPDLKRPDLATYEAIRKSGIMPTVLANANEIYAEAFINEEYKFTDIAVI 355
294 CKLGSILFKKPDNVKYPSPMDLAVAAAGAGTMTGVLISANEKAVEMFIDEKISYDIPKV 353
356 IERSMAQFKPD--DAGSLELVLAQDQDAREVARDI 388
354 VELTCDKHRELVTSPSLEIIVHYDLMAREYANV 388

RESULT 5

US-09-449-335-6
Sequence 6, Application US/09449335
Patent No. 6303365
GENERAL INFORMATION:
APPLICANT: Busch, Marco
APPLICANT: Hain, Rudiger
APPLICANT: Martin, William
APPLICANT: Tietjen, Klaus
APPLICANT: Klotz, Andreas
TITLE OF INVENTION: Method of determining the activity of
TITLE OF INVENTION: 1-deoxy-D-xylinose-5-phosphate reductoisomerase and
FILE REFERENCE: 202005
CURRENT APPLICATION NUMBER: US/09/449,335
CURRENT FILING DATE: 1999-11-24
EARLIER APPLICATION NUMBER: DE 199 35 967.9
EARLIER FILING DATE: 1999-07-30
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 6
LENGTH: 477
TYPE: PRF
ORGANISM: Arabidopsis thaliana
US-09-449-335-6

Query Match

41.2%; Score 818.5; DB 4; Length 477;
Best Local Similarity 44.6%; Pred. No. 2,2e-81;
Matches 176; Conservative 71; Mismatches 131; Indels 17; Gaps 5;

2 KCICILGATGSGIVSTLDVVAHPRKQVVALTANGINDALYEOCLAHPRVAVVWESK 61
81 KRISIVSGTSGIGTDLIVAENPKFRVVALAAGSVTLADQVRRPKPALVAVRNSL 140
62 VAEFKRIASPVADI---KVLSSGSEALQOVATLENVDTMAIYGAAGLPTTAAANAGKTV 121
141 INELKEALADLVY-KLEIIPGEGVIEVARHPEAVTVTGICAGLKPVTAAIEAGDI 199
122 LLANKEALVMSGQIFMGAVSDGAVLLPIDSEHNAIFOC---MPAGYTPGHTAKQARRIL 178
200 ALANKETLIAGSPFLPLANKKNVILIPADSEHNAIFOCIOGLPBG-----ALRKII 251
179 LTASGAPRRRTIEFLSSVTPDOAVAHPRKMDGRIKISVDSATMNGKIELLEIACILFPMME 238
252 LTASGAPRRRTIEFLSSVTPDOAVAHPRKMDGRIKISVDSATMNGKIELLEIACILFPMME 233
239 PPOIEVVIHPOSIHSMVDYVDSVLAQMGNDPMTPIAHAMAPRRPDSGVAP---LDI 295
312 YDDIEIVIHPOSIHSMVDYVDSVLAQMGNDPMTPIAHAMAPRRPDSGVAP---LDI 293
236 FEVGNDEPKPDLKRPDLATYEAIRKSGIMPTVLANANEIYAEAFINEEYKFTDIAVI 355

372 CKLGSILFKKPDNVKYPSPMDLAVAAAGAGTMTGVLISANEKAVEMFIDEKISYDIPKV 431
356 IERSMAQFKPD--DAGSLELVLAQDQDAREVARDI 388
432 VELTCDKHRELVTSPSLEIIVHYDLMAREYANV 466

RESULT 6

US-09-480-921B-10
Sequence 10, Application US/09480921B
Patent No. 6387637
GENERAL INFORMATION:
APPLICANT: Levin, Joshua Z.
APPLICANT: Budziszewski, Gregory J.
APPLICANT: Potter, Sharon L.
APPLICANT: Megrich, Lynette M.
TITLE OF INVENTION: Herbicide Target Genes and Methods
FILE REFERENCE: PB/5-30780A
CURRENT APPLICATION NUMBER: US/09/480,921B
CURRENT FILING DATE: 2000-01-11
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 10
LENGTH: 477
TYPE: PRF
ORGANISM: Arabidopsis thaliana
US-09-480-921B-10

Query Match 41.2%; Score 817.5; DB 4; Length 477;
Best Local Similarity 44.6%; Pred. No. 2,2e-81;
Matches 178; Conservative 69; Mismatches 127; Indels 25; Gaps 6;

2 KCICILGATGSGIVSTLDVVAHPRKQVVALTANGINDALYEOCLAHPRVAVVWESK 61
81 KRISIVSGTSGIGTDLIVAENPKFRVVALAAGSVTLADQVRRPKPALVAVRNSL 140
62 VAEFKRIASPVADI---KVLSSGSEALQOVATLENVDTMAIYGAAGLPTTAAANAGKTV 117
141 INELKEALADLVY-KLEIIPGEGVIEVARHPEAVTVTGICAGLKPVTAAIEAGDI 195
118 GKTVLANKEALVMSGQIFMGAVSDGAVLLPIDSEHNAIFOC---MPAGYTPGHTAKQARRIL 174
196 GKTVLANKEALVMSGQIFMGAVSDGAVLLPIDSEHNAIFOCIOGLPBG-----ALRKII 247
175 KRILITASGAPRRRTIEFLSSVTPDOAVAHPRKMDGRIKISVDSATMNGKIELLEIACILFPMME 234
248 KRILITASGAPRRRTIEFLSSVTPDOAVAHPRKMDGRIKISVDSATMNGKIELLEIACILFPMME 233
235 FMEPPDIEVVIHPOSIHSMVDYVDSVLAQMGNDPMTPIAHAMAPRRPDSGVAP---LDI 292
308 FGAEYDDIEIVIHPOSIHSMVDYVDSVLAQMGNDPMTPIAHAMAPRRPDSGVAP---LDI 293
293 -LDI-FEVGNDEPKPDLKRPDLATYEAIRKSGIMPTVLANANEIYAEAFINEEYKFTDIAVI 351
368 RLDLCKLGSILFKKPDNVKYPSPMDLAVAAAGAGTMTGVLISANEKAVEMFIDEKISYDIPKV 427
352 IAVIERSMAQFKPD--DAGSLELVLAQDQDAREVARDI 388
428 IFKVELTCDKHRELVTSPSLEIIVHYDLMAREYANV 466

RESULT 7

US-09-449-335-2
Sequence 2, Application US/09449335
Patent No. 6303365
GENERAL INFORMATION:
APPLICANT: Busch, Marco
APPLICANT: Hain, Rudiger
APPLICANT: Martin, William
APPLICANT: Tietjen, Klaus
APPLICANT: Klotz, Andreas
TITLE OF INVENTION: Method of determining the activity of

TITLE OF INVENTION: 1-deoxy-D-xylulose-5-phosphate reductoisomerase and
 TITLE OF INVENTION: 1-deoxy-D-xylulose-5-phosphate synthase
 FILE REFERENCE: 2020US
 CURRENT APPLICATION NUMBER: US/09/449,335
 CURRENT FILING DATE: 1999-11-24
 EARLIER APPLICATION NUMBER: DB 199 35 967.9
 EARLIER FILING DATE: 1999-07-30
 NUMBER OF SEQ ID NOS: 8
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2
 LENGTH: 477
 TYPE: PRT
 ORGANISM: Arabidopsis thaliana
 US-09-449-335-2

Query Match 41.0%; Score 814.5; DB 4; Length 477;
 Best Local Similarity 44.3%; Pred. No. 6.1e-81;
 Matches 175; Conservative 71; Mismatches 132; Indels 17; Gaps 5;

QY 2 KGCILGATGSGVSTLDVVAHHPKQVVALTANGIDALYEQCLAHPEYAVVWESK 61
 DB 81 KPISIVGSGSIGTQTLDIVAENPKFRVVALAAGSVTLADQVRRFRKPAVAVRNSL 140
 QY 62 VAEFKRIASPVADI---KVLGSEALQOVATLENVDTMAAIVGAGLPTLAAAKGTV 121
 DB 141 INELKEALADLY-KLEIIRGEGVIEVARHPEAVTVTGIVGAGLPTVAALIEAGDI 199
 QY 122 LIANKEALVMSQIIMQAVSDGAVLPIDSEHNAIFOC---MPAGYTPGHTAKQARIL 178
 DB 200 ALANKETILAGPFVPLANKKNVXKLLPADSEHSAIFOCIQGLPEG-----ALRKII 251
 QY 179 LTAGSGPFRPTPIETLSSVTPDOVAHHPKQVVALTANGIDALYEQCLAHPEYAVVWESK 238
 DB 252 LTAGSGAFRDMVEKLEKVKADALKHPNMNKKITVDSCTLPKGLVIRAHILFQAE 311
 QY 239 PDQIEVVIHPQSIHSMVDYVDSVLAQMGNDPKPTIAHAAWBERFDSGVAP---LDI 295
 DB 312 YDIEIIVHPQSIHSMITQDSVLAQMGNDPKPTIYMSMDRVPCEVTPRDL 371
 QY 296 FEVGMDEPKPKPCPCRLAYEALKSGIMPTVLAANEIAVEAFINEVKFTDIAVI 355
 DB 372 CKLGSILTFKKPDVVKPSPMDLAVAAAGAGTGTGVALANKEAVEMFIDKISYLDIKV 431
 QY 356 IERSMAQFKPD--DAGSLVELVQADQDAFEVARDI 388
 DB 432 VZLTCDKHRELVTSPSLEIIVHYDLMAREYAAV 466

RESULT 8
 US-09-480-921B-29
 Sequence 29, Application US/09480921B
 Patent No. 6387637
 GENERAL INFORMATION:
 APPLICANT: Levitz, Joshua Z.
 APPLICANT: Buddiszewski, Gregory J.
 APPLICANT: Foster, Sharon L.
 APPLICANT: Weglich, Lynette M.
 TITLE OF INVENTION: Herbicide Target Genes and Methods
 FILE REFERENCE: PB/5-30780A
 CURRENT APPLICATION NUMBER: US/09/480,921B
 CURRENT FILING DATE: 2000-01-11
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 29
 LENGTH: 477
 TYPE: PRT
 ORGANISM: Arabidopsis thaliana
 FEATURE:
 NAME/KEY: SITE
 LOCATION: (39)
 OTHER INFORMATION: Xaa = Asp or Asn
 NAME/KEY: SITE
 LOCATION: (155)

OTHER INFORMATION: Xaa = Pro or Leu
 NAME/KEY: SITE
 LOCATION: (162)
 OTHER INFORMATION: Xaa = Leu or Gln
 NAME/KEY: SITE
 LOCATION: (187)
 OTHER INFORMATION: Xaa = Lys or Gln
 NAME/KEY: SITE
 LOCATION: (465)
 OTHER INFORMATION: Xaa = Asp or Asn
 US-09-480-921B-29

Query Match 41.0%; Score 813.5; DB 4; Length 477;
 Best Local Similarity 44.4%; Pred. No. 7.8e-81;
 Matches 177; Conservative 69; Mismatches 128; Indels 25; Gaps 6;

QY 2 KGCILGATGSGVSTLDVVAHHPKQVVALTANGIDALYEQCLAHPEYAVVWESK 61
 DB 81 KPISIVGSGSIGTQTLDIVAENPKFRVVALAAGSVTLADQVRRFRKPAVAVRNSL 140
 QY 62 VAEFKRIASPVADI---KVLGSEALQOVATLENVDTMAAIVGAGLPTLAAAKA 117
 DB 141 INELKEAL-----ADLYKXELIIPGEGVIEVARHPEAVTVTGIVGAGLPTVAALIEA 195
 QY 118 GKVTLANKALVMSQIIMQAVSDGAVLPIDSEHNAIFOC---MPAGYTPGHTAKQA 174
 DB 196 GKVTLANKETILAGPFVPLANKKNVXKLLPADSEHSAIFOCIQGLPEG-----AL 247
 QY 175 RRLTLASGPFRRPTPIETLSSVTPDOVAHHPKQVVALTANGIDALYEQCLAHPEYAVVWESK 234
 DB 248 RKILITLSSGAFRDMVEKLEKVKADALKHPNMNKKITVDSCTLPKGLVIRAHILFQAE 307
 QY 235 FNMEDQIEVVIHPQSIHSMVDYVDSVLAQMGNDPKPTIAHAAWBERFDSGVAP---LDI 292
 DB 308 FGAIEDIIBVHPQSIHSMITQDSVLAQMGNDPKPTIYMSMDRVPCEVTPRDL 367
 QY 293 LDIIEVGMDEPKPKPCPCRLAYEALKSGIMPTVLAANEIAVEAFINEVKFTD 351
 DB 368 RLDLCKLGSILTFKKPDVVKPSPMDLAVAAAGAGTGTGVALANKEAVEMFIDKISYLD 427
 QY 352 IAVIERSMAQFKPD--DAGSLVELVQADQDAFEVARDI 388
 DB 428 IFKVELTCDKHRELVTSPSLEIIVHYDLMAREYAAV 466

RESULT 9
 US-09-491-362-2
 Sequence 2, Application US/09491362
 Patent No. 6281017
 GENERAL INFORMATION:
 APPLICANT: Croteau, Rodney B
 APPLICANT: Lange, Bernd M
 TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
 TITLE OF INVENTION: METHODS OF USE
 FILE REFERENCE: NSUR14972
 CURRENT APPLICATION NUMBER: US/09/491,362
 CURRENT FILING DATE: 2000-01-26
 EARLIER APPLICATION NUMBER: 60/118,349
 EARLIER FILING DATE: 1999-02-03
 NUMBER OF SEQ ID NOS: 13
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 2
 LENGTH: 475
 TYPE: PRT
 ORGANISM: Mentha piperita
 US-09-491-362-2

Query Match 37.3%; Score 741; DB 3; Length 475;
 Best Local Similarity 43.0%; Pred. No. 7.8e-73;
 Matches 172; Conservative 73; Mismatches 133; Indels 22; Gaps 10;
 QY 2 KGCILGATGSGVSTLDVVAHHPKQVVALTANGIDALYEQCLAHPEYAVVWESK 61
 DB 81 KPISIVGSGSIGTQTLDIVAENPKFRVVALAAGSVTLADQVRRFRKPAVAVRNSL 140

Db 80 KPIVIGTSGTGTOTLDIVANPDKFRIVALAAGNTVLADQ-KAKPKLVSVKDESL 138
 QY 62 VAEFKORIAASPVADI-KVLSGSEALQQVATLENVDTWMAIVGAAGLLPTLAAAKAGT 120
 Db 139 ISELKEALAG--FEDMPEIIPGEQGLI EVARPDVAIVTGTIVGAGLKPFTVAIEAGD 196
 QY 121 VILAKKEALVNSGOIFMOAVSDSGAVLLPIDSEHNAIFQC---MAGTIPGHTAQARI 177
 Db 197 IALAKKEALVNSGOIFMOAVSDSGAVLLPIDSEHNAIFQC---MAGTIPGHTAQARI 248
 QY 178 ILTASGPFRTPIETLSSVTPDOAVAHPKMDGKRIISVDSATM-MNKGLEIIEACLLFN 236
 Db 249 ILTASGPFRTPIETLSSVTPDOAVAHPKMDGKRIISVDSATM-MNKGLEIIEACLLFN 308
 QY 237 MEPPDIEVIVHPSQIISHWVDVDSVLAQMGKPMKRIPIAHAMWPER-FDSGVA--PL 293
 Db 309 AEPDIEIVHPSQIISHWVDVDSVLAQMGKPMKRIPIAHAMWPER-FDSGVA--PL 368
 QY 294 DIFEGHMDPEKPDLPKPCRLAYEAISGGINPTVLANAEIIVAEALINEVKTIDIA 353
 Db 369 DLCKV-DLPEKPDNREIPAMDIAAAMKSRSTGTGVLISANERKAVEMFIDEKIGYDIF 427
 QY 354 VILERSMAQFKPDDA--GSLELVLAQADQAREVARDIIRT 391
 Db 428 KVEELTCDKRESEMAVSPSLEIIVHYDQWADYATVLS 467

RESULT 10

US-09-874-562-2
 / Sequence 2, Application US/09674562
 / Patent No. 6420159
 / GENERAL INFORMATION:
 / APPLICANT: Croteau, Rodney B
 / TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
 / FILE REFERENCE: MSU17549
 / CURRENT APPLICATION NUMBER: US/09/874,562
 / PRIOR FILING DATE: 2001-06-04
 / PRIOR APPLICATION NUMBER: 09/491,362
 / PRIOR FILING DATE: 1999-02-03
 / NUMBER OF SEQ ID NOS: 13
 / SOFTWARE: Patent In Ver. 2.0
 / SEQ ID NO 2
 / LENGTH: 475
 / TYPE: PRT
 / ORGANISM: Mentha piperita
 / US-09-874-562-2

Query Match 37.3%; Score 741; DB 4; Length 475;
 Best Local Similarity 43.0%; Pred. No. 7,8e-73;
 Matches 172; Conservative 73; Mismatches 133; Indels 22; Gaps 10;

QY 2 KICILGATGSGTLDIVANPDKFRIVALAAGNTVLADQ-KAKPKLVSVKDESL 61
 Db 80 KPIVIGTSGTGTOTLDIVANPDKFRIVALAAGNTVLADQ-KAKPKLVSVKDESL 138
 QY 62 VAEFKORIAASPVADI-KVLSGSEALQQVATLENVDTWMAIVGAAGLLPTLAAAKAGT 120
 Db 139 ISELKEALAG--FEDMPEIIPGEQGLI EVARPDVAIVTGTIVGAGLKPFTVAIEAGD 196
 QY 121 VILAKKEALVNSGOIFMOAVSDSGAVLLPIDSEHNAIFQC---MAGTIPGHTAQARI 177
 Db 197 IALAKKEALVNSGOIFMOAVSDSGAVLLPIDSEHNAIFQC---MAGTIPGHTAQARI 248
 QY 178 ILTASGPFRTPIETLSSVTPDOAVAHPKMDGKRIISVDSATM-MNKGLEIIEACLLFN 236
 Db 249 ILTASGPFRTPIETLSSVTPDOAVAHPKMDGKRIISVDSATM-MNKGLEIIEACLLFN 308
 QY 237 MEPPDIEVIVHPSQIISHWVDVDSVLAQMGKPMKRIPIAHAMWPER-FDSGVA--PL 293
 Db 309 AEPDIEIVHPSQIISHWVDVDSVLAQMGKPMKRIPIAHAMWPER-FDSGVA--PL 368

Db 309 AEPDIEIVHPSQIISHWVDVDSVLAQMGKPMKRIPIAHAMWPER-FDSGVA--PL 368
 QY 294 DIFEGHMDPEKPDLPKPCRLAYEAISGGINPTVLANAEIIVAEALINEVKTIDIA 353
 Db 369 DLCKV-DLPEKPDNREIPAMDIAAAMKSRSTGTGVLISANERKAVEMFIDEKIGYDIF 427
 QY 354 VILERSMAQFKPDDA--GSLELVLAQADQAREVARDIIRT 391
 Db 428 KVEELTCDKRESEMAVSPSLEIIVHYDQWADYATVLS 467

RESULT 11

US-09-198-452A-362
 / Sequence 362, Application US/09198452A
 / Patent No. 6559294
 / GENERAL INFORMATION:
 / APPLICANT: Griffiths, R.
 / TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 / thereof and uses thereof, in particular for the diagnosis, prevention
 / and treatment of infection
 / FILE REFERENCE: 9710-003-999
 / CURRENT APPLICATION NUMBER: US/09/198,452A
 / PRIOR FILING DATE: 1998-11-24
 / NUMBER OF SEQ ID NOS: 6849
 / SEQ ID NO 362
 / LENGTH: 340
 / TYPE: PRT
 / ORGANISM: Chlamydia pneumoniae
 / US-09-198-452A-362

Query Match 31.2%; Score 620; DB 4; Length 340;
 Best Local Similarity 39.8%; Pred. No. 1e-59;
 Matches 138; Conservative 65; Mismatches 134; Indels 10; Gaps 2;

QY 42 LYECGLAHPEYAVYVNSKAEKORIAASPVADIVTSGSEALQQVATLENVDTWMA 101
 Db 4 IFQGLQEPAPLAAVYVEVNEACQRF-----PHQFPGQEBLTQICMDVTTVAA 58
 QY 102 IVGAAGLLPTLAAAKAGTIVLANKEALVNSGOIFMOAVSDSGAVLLPIDSEHNAIFQC 161
 Db 59 SSGLEALPILLESKKKALANKELIVCAGELVSTAKENGKVLPISEHNAIFQC 118
 QY 162 PAGTIPGHTAQARIILTLASGPFRTPIETLSSVTPDOAVAHPKMDGKRIISVDSATM 221
 Db 119 -----EGRTLEGIKELTLASGPFRTPIETLSSVTPDOAVAHPKMDGKRIISVDSATM 173
 QY 222 MNKGLEIIEACLLFNMEPPDIEVIVHPSQIISHWVDVDSVLAQMGKPMKRIPIAHAM 281
 Db 174 VNKGLEIIEACLLFNMEPPDIEVIVHPSQIISHWVDVDSVLAQMGKPMKRIPIAHAM 233
 QY 282 WPERFDSGVAPLDIFEGHMDPEKPDLPKPCRLAYEAISGGINPTVLANAEIIVAE 341
 Db 234 ABERFASPRDQMDPSKKQTLFPVDERPPIRLAQVLEKQSSGSPFAAAEVLVR 293
 QY 342 PLNEVKTIDIAVILERSMAQFKPDDAGSLELVLAQADQAREVARDIIRT 388
 Db 294 FLECEISWCDILKRLTITMCHKAYACHSLEDILVGEAPALAOET 340

RESULT 12

US-09-252-991A-31862
 / Sequence 31862, Application US/09252991A
 / Patent No. 6551795
 / GENERAL INFORMATION:
 / APPLICANT: Marc J. Rubenfield et al.
 / TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 / FILE REFERENCE: 107196, 136
 / CURRENT APPLICATION NUMBER: US/09/252,991A
 / PRIOR FILING DATE: 1999-02-18
 / PRIOR APPLICATION NUMBER: US 60/074,788
 / PRIOR FILING DATE: 1998-02-18
 / PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 31862
 LENGTH: 671
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-31862

Query Match
 Best Local Similarity 21.1%; Score 111; DB 4; Length 671;
 Matches 85; Conservative 58; Mismatches 139; Indels 120; Gaps 20;

QY 9 ATGSGVSTLDVVAHPDKYQVVALTANGNIDALVEQ-----CLAHPEYAVVWMS--K 61
 DB 302 AAGLVVAARLDLAAH-----AGCDLHIIIEERQPVGQRAAEVGFGLNAAVE 351
 QY 62 VAEFKRIASPVADIKVLGSGEALQO-----VATLENDVTMAAIVGAAAG 107
 DB 352 AAEVHQRRRRRGVLEVARLHPGEGIDHAPFVQAVGAERPVYAEERAGVAAQADLGIG 411
 QY 108 -LLPTLAARKAKTVLLANKKALVMSGOIFMOAVSDSG-----AVLLPDSHNAIF 158
 DB 412 AAVALAACQAVDDIALVADVEDAVGDFVVALAADAGERRQVGEVLAAND--HGVL 469
 QY 159 -----QCM-----PAGYTPGHTAQARRIILITASGPPRRPTIEFTLSV 197
 DB 470 LVQVGIQAACACLFKRMGVDGSGRAGQORAHHAFOAFILVQVAG-----QAKLA 521
 QY 198 TPDOAVA-HPKV-----DMGRKISVDSATMNMKGLEIACLLFMPEPDQIEV-----VI 246
 DB 522 TEVQAGHSGKLAALMTDGGIADVADVAHAAGVLAER-----PDVEVSALAL 574
 QY 247 HP-----QSIHNSVDVYDSV-LAQMNPDKETPIAHAMAMBERPDSGVAPLDIEF- 297
 DB 575 GPIDADAGQLVYGAAGHQVDDAADAAGNP-----AEQVWMSLEHFA 619
 QY 298 VEHMDEKEDLGRFCLRLAYEAKISGGIMPTVLANAEIAV 339
 DB 620 LDEFDPYPERQO-----AIEIWE-GDVGVLAHAERAAVAV 654

RESULT 13
 US-09-252-991A-17790
 Sequence 17790, Application US/09252991A
 Patent No. 6551735
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 17790
 LENGTH: 2285
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-17790

Query Match
 Best Local Similarity 22.8%; Pred. No. 0.11;
 Matches 87; Conservative 51; Mismatches 137; Indels 106; Gaps 19;

QY 6 ILGATGSGVSTLDVVAHPDKYQVVALTANGNIDALVEQCLAHPEYAVVWMSKVAEF 65
 DB 1505 LIGAT--FVELLMRLAEHPDRAADFAFLAPRR-DA-----ASQPEPLVDV---VALF 1550
 QY 66 KORIASPVADIKVLGSGEAL--CQVATLENDVTMAAIVGAAAGLLPTLAARKAKTVL 122

DB 1551 ERQVEALP-----GSAALAPBEQRWTRYDLDHVARCV--ATRLV--PAGARRGDAIG 1598
 QY 123 LANKKALVMSGOIFMOAVSDSGAVLLPDSHNAIFQCMFAGTTPGHTAQARRIILITAS 182
 DB 1599 VALNRSPREMIATVV--GILRAGIACVPLDIVSYPA-----QRTALLLET 1639
 QY 183 GGPFRRT-----PIETLSVTPDOAVAHPKMD-MGRKISVDSATMNMKG 226
 DB 1640 AQPFRVVAHPENHAVALAAERVLVPEELVADIKPETFAPQDELAMLLFTSGSGRRKGV 1659
 QY 227 ELIEACLLFNNPEPDQIEVVIHPOSITHSM--VDVYDGSVLAQMGNPDKRT---PIAHAMA 281
 DB 1700 EL-----SHRMANYTQMQLRYASGVPGILRTIQAPLSPDMA 1736
 QY 282 WPERFDS--GVAPLD--IEVGHMD-----FEKPDLR---FPCLRLEYAKISGGIM 327
 DB 1737 FOELFSTLCGGEGEQLQILSNRRERDPSALLHYLEERQVRLPVALQRLAENSALGV 1796
 QY 328 PTVLANAEIAVEAFINERYK 348
 DB 1797 FGALRVVVSSEQLRTIEDVR 1817

RESULT 14
 US-09-428-517-3
 Sequence 3, Application US/09428517
 Patent No. 6251636
 GENERAL INFORMATION:
 APPLICANT: Betlach, Mary C.
 APPLICANT: Shah, Sanjay Krishnakant
 APPLICANT: McDaniel, Robert
 TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
 FILE REFERENCE: 30062-20029.00
 CURRENT APPLICATION NUMBER: US/09/428,517
 CURRENT FILING DATE: 1999-10-28
 EARLIER APPLICATION NUMBER: 60/120,254
 EARLIER FILING DATE: 1999-02-16
 EARLIER APPLICATION NUMBER: 60/106,100
 EARLIER FILING DATE: 1998-10-29
 NUMBER OF SEQ ID NOS: 12
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 3
 LENGTH: 3816
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Recombinant
 OTHER INFORMATION: Oleandolide PKS
 US-09-428-517-3

Query Match
 Best Local Similarity 17.8%; Pred. No. 0.41;
 Matches 98; Conservative 76; Mismatches 151; Indels 227; Gaps 25;

QY 2 KGIICLAFGSGVSTLDVVAHPDKYQVVALTANGNIDALVEQCLAHPEYAVVWMSK 61
 DB 3163 ESTVYHSAAGVGMAAVQ-LARKMDA-EVFGTAKSGKMDVLAAGLDBEH-----IGSSR 3215
 QY 62 VAEFKRIASPVADIKVLGSGEALQVATLENDVTMAA----- 101
 DB 3216 TTEFEGRFRAT-----SGSGRIDVAVNALSGDFDASARLIREGRFVEMKTDIRT 3267
 QY 102 ---IVGAAGLLPTLA-----AAKAGKTVLLANKKALVMSGOIF--MOAVSDSGAVLL-- 148
 DB 3268 DLGVVAGDG--VPDIRYAFDLAENGARF-----CQMLDEIYALTDAGTILRPL 3316
 QY 149 ---PIDSEHNAIFQCMFAGTTPGHTAQAR-- 183
 DB 3317 RAMVVRARHAL-----RFVSAHRHVGVVLTVPALDABGTVALITGAGTICAL 3365
 QY 184 -----GPPRRPTIEFTLSVTPDOAVAHPKMDMKKISVDSATMNM--KGLS----- 227

Db 3366 VARRLVTHDVFRLLVSRSGVAPD-LAAELGALCAEYTAACDVANRKAALKALBDIP 3423
QY 228 -----LIEACILFNMEDQIEVYTHPQ-----SIHSMVYVD-----G 261
Db 3424 PEHPVTVGVHAGVLDGVDGVSGLTPERVDTYVKKPVDAALTLIESITGSLDDPALFVIFS 3483
QY 262 SYLAQMGNDP-----MRTPIAH-----AMN-----PER 285
Db 3484 SAASKLGGGGGQSYAANQFIDTLARHARRGLTSVLSGMGMHEBASGLTGGADIDRDR 3543
QY 286 FD-SGVAPLDFEVGEM-----DFEKPDLKRPFCRLAYEAIKSGGIMPTVYN----- 332
Db 3544 MSRAQIAIMPIDRALHLPDRATELGDVPLIPRLNBALEDBRAADGTPPLISGLVRYRH 3603
QY 333 -----AANETAVEAPLNEEYKFTDIAVIERSMAGFKPDAGSLLEVLAQDQDAR 382
Db 3604 RPSAPAGTATAAPATGPEAFARE-----LAAAPDPR 3634
QY 383 EVARDIITLVVA 394
Db 3635 RALRDLVGRHVA 3646

RESULT 15
US-09-252-991A-20932
; Sequence 20932, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20932
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20932

Query Match 5.0%; Score 99; DB 4; Length 482;

Best Local Similarity 20.3%; Pred. No. 0.059; Indels 124; Gaps 18;
Matches 86; Conservative 57; Mismatches 156;

QY 20 VVAHPDKYQVVALTANG-NIDALYEQCLAHPEYAVVWESKYA-----EPKORI 69
Db 50 VVRQGESVEALGLAAGFGDLAGDVGVHAAVAAITGIVDAADADLGAREGGQEV 109
QY 70 AASPVAADIKVLSGSEALQCVATLENVDT--VMAIVGAAGLPTLLAAKAGKTVLLANKE 127
Db 110 AGPGVDLHVQLREGLFHRPDRLDVRITSAIDHATIDQALVAG-----B 158
QY 128 ALVNSGGCIFMGAVSDSGAVLPIIDSEHNAIFQCMPPAGTTPGHTAKQARRILITASGPF 187
Db 159 ALIVE-----QVIAVFAHVLT-----RQAVGQFLVQRLGSDHLGAAGHRL-----GROFR 204
QY 188 RPIETLTSSVTPDQAVAH-----PKMDGRKISVDSATMNMKGLELEACILFN 236
Db 205 DOAVEVGTGSHDELRLIALRGHHRAGALDAGR-----ALLVNSAERLHRRRPAE 259
QY 237 MEPOIEV-----VIHPGSI--HSMVYVDGSVLAQWGNPMRTPIAHAMWPERFDSG 289
Db 260 GEVQGMVDAAHVKAHTIIVGRHHD-----ALGVHDLDLV--NVAFPQLFLRA 309
QY 290 VAPLDIFEVGMDEPKDLKFPCLRLAYEAIKSGGIMPTVYNANET-----A 338
Db 310 -----QVIRH-----LIGGGEHPAVLQVALDPTVAGHALADADA 344

QY 339 VEAPLNEEYKFTDIAVIERSMAGFKPDDA-----GSIEVLVLAQDQD 380
Db 345 FEGHIAEQLEFALADGALDHYDVAAVAVDILAAVAPGSAEDLGSGFGYGLLETILQOEG 404
QY 381 ARE 383
Db 405 GGE 407

Search completed: January 29, 2004, 15:57:03
Job time : 13.3656 secs

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

Run on: January 29, 2004, 15:54:19 ; Search time 26.8993 Seconds
(without alignments)
3044.503 Million cell updates/sec

Title: US-09-941-947a-8

Perfect score: 1985
Sequence: 1 MKGICILGATSGISVSTLDV.....LQADQARVARDIKITVA 394

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1985	100.0	394	10	US-09-934-903-4
2	1985	100.0	394	10	US-09-934-868-64
3	1985	100.0	394	11	US-09-941-947a-8
4	1215.5	61.2	396	12	US-10-381-779-117
5	1052.5	53.0	394	12	US-10-381-779-119
6	1045	52.6	398	12	US-10-381-779-100
7	1045	52.6	398	12	US-10-381-779-118
8	1042	52.5	405	12	US-10-381-779-122
9	1007	50.7	397	12	US-10-381-779-99
10	1007	50.7	397	12	US-10-381-779-120
11	999.5	50.4	388	12	US-10-381-779-101
12	999.5	50.4	388	12	US-10-381-779-116
13	994.5	45.0	386	12	US-10-381-779-97
14	892.5	45.0	394	12	US-10-381-779-103
15	892.5	45.0	394	12	US-10-381-779-121

16	886	44.6	402	12	US-10-381-779-102	Sequence 102, App
17	886	44.6	402	12	US-10-381-779-123	Sequence 123, App
18	818.5	41.2	477	9	US-09-923-556-6	Sequence 6, Appl
19	818.5	41.2	477	10	US-09-987-025-2	Sequence 2, Appl
20	818.5	41.2	477	12	US-10-381-779-128	Sequence 128, App
21	817.5	41.0	477	14	US-10-047-412a-10	Sequence 10, Appl
22	814.5	41.0	477	9	US-09-923-556-2	Sequence 2, Appl
23	813.5	41.0	477	14	US-10-047-412a-29	Sequence 29, Appl
24	798.5	40.2	388	12	US-10-381-779-98	Sequence 98, Appl
25	798.5	40.2	388	12	US-10-381-779-125	Sequence 125, App
26	731.5	36.9	385	12	US-10-128-713a-4	Sequence 4, Appl
27	729.5	36.8	486	12	US-10-381-779-131	Sequence 131, App
28	721.5	36.3	487	12	US-10-259-194a-268	Sequence 268, App
29	718	36.2	386	12	US-10-381-779-124	Sequence 124, App
30	697	35.1	394	15	US-10-156-761-10100	Sequence 10100, A
31	692	34.9	392	10	US-09-738-626-5709	Sequence 5709, App
32	691	34.8	436	10	US-09-712-363-258	Sequence 258, App
33	691	34.8	436	12	US-10-381-779-104	Sequence 104, App
34	691	34.8	436	12	US-10-381-779-127	Sequence 127, App
35	681	34.3	406	12	US-10-381-779-126	Sequence 126, App
36	620	31.2	340	12	US-10-289-362-362	Sequence 362, App
37	596	30.0	1588	12	US-09-925-778-2	Sequence 2, Appl
38	486	24.5	356	12	US-10-381-779-130	Sequence 130, App
39	473	23.8	368	12	US-09-882-227-500	Sequence 500, Appl
40	428.5	21.6	177	12	US-10-032-395-1	Sequence 1, Appl
41	251	12.6	98	12	US-09-864-408a-8646	Sequence 8646, App
42	114	5.7	425	12	US-10-369-493-21773	Sequence 21773, A
43	113.5	5.7	1073	15	US-10-156-761-12156	Sequence 12156, A
44	109.5	5.5	6146	15	US-10-156-761-10436	Sequence 10436, A
45	109	5.5	426	12	US-10-369-493-21804	Sequence 21804, A

ALIGNMENTS

RESULT 1
US-09-934-903-4
Sequence 4, Application US/09934903
Patent No. US20020102690A1
GENERAL INFORMATION:
APPLICANT: Koffas, Matthews
APPLICANT: Odum, J. Martin
APPLICANT: Schenzle, Andreas J.
APPLICANT: No. US20020102690A1ton, Kelley C.
APPLICANT: Tomb, Jean-Francois
APPLICANT: Rouviere, Pierre
APPLICANT: Picotagisic, Stephen
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
FILE REFERENCE: C11646 US NA
CURRENT APPLICATION NUMBER: US/09/334, 903
CURRENT FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/229,907
PRIOR FILING DATE: September 1, 2001
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 4
LENGTH: 394
TYPE: PRT
ORGANISM: Methylobionas 16a
FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by ORF2
US-09-934-903-4
Query Match 100.0%; Score 1985; DB 10; Length 394;
Best Local Similarity 100.0%; Pred. No. 7.2e-186;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKGICILGATSGISVSTLDVARRHDPKQVAVLRANGIDALVYOCCLAHREYVWVWES 60
DB 1 MKGICILGATSGISVSTLDVARRHDPKQVAVLRANGIDALVYOCCLAHREYVWVWES 60
QY 61 KVAEFKRIAPASPVADIKVLSGSEALQVATLENTVMAIVGAAGILPTLAARAKAKGT 120

```

Db      61 KVAEFKQRIASPVAADIKVLSGSEALQOVATLENDVTMAAIVGAAGLLPTLAAAKAGKT 120
      121 VILANKELVMSGOIFMOWSDSGAVLLPISSEHNAIQCPAGTTPGHTAKOARRILLT 180
      121 VILANKELVMSGOIFMOWSDSGAVLLPISSEHNAIQCPAGTTPGHTAKOARRILLT 180
      121 VILANKELVMSGOIFMOWSDSGAVLLPISSEHNAIQCPAGTTPGHTAKOARRILLT 180
      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVDSATMANKGLIELIACILFNMEPD 240
      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVDSATMANKGLIELIACILFNMEPD 240
      241 QIEVVIHQSIHSMVDYVDSVLAQMGKPMRTPIAHAMAMPFDSGVAPLDIFEVGH 300
      241 QIEVVIHQSIHSMVDYVDSVLAQMGKPMRTPIAHAMAMPFDSGVAPLDIFEVGH 300
      301 MDEFKPDLRFPCLRLAYAIKSGGIMPTVLANANEIAVEAFINEEVFTDIAVIERSM 360
      301 MDEFKPDLRFPCLRLAYAIKSGGIMPTVLANANEIAVEAFINEEVFTDIAVIERSM 360
      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394
      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394

```

RESULT 2

```

US-09-934-868-64
; Sequence 64, Application US/09934868
; Patent No. US20020137190A1
; GENERAL INFORMATION:
; APPLICANT: Koffas, Matchee
; APPLICANT: Odom, James M.
; APPLICANT: Schenck, Andreas J.
; TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN
; FILE REFERENCE: CL1596 US NA
; CURRENT APPLICATION NUMBER: US/09/934,868
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 64
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Methylobionas 16a
; FEATURE:
; OTHER INFORMATION: Amino acid sequences encoded by DXR
US-09-934-868-64

```

```

Query Match      100.0%; Score 1985; DB 10; Length 394;
Best Local Similarity 100.0%; Pred. No. 7,2e-186;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

      1 MKGICILGATGSIQVSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHREYAVVWMS 60
      1 MKGICILGATGSIQVSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHREYAVVWMS 60
      61 KVAEFKQRIASPVAADIKVLSGSEALQOVATLENDVTMAAIVGAAGLLPTLAAAKAGKT 120
      61 KVAEFKQRIASPVAADIKVLSGSEALQOVATLENDVTMAAIVGAAGLLPTLAAAKAGKT 120
      121 VILANKELVMSGOIFMOWSDSGAVLLPISSEHNAIQCPAGTTPGHTAKOARRILLT 180
      121 VILANKELVMSGOIFMOWSDSGAVLLPISSEHNAIQCPAGTTPGHTAKOARRILLT 180
      121 VILANKELVMSGOIFMOWSDSGAVLLPISSEHNAIQCPAGTTPGHTAKOARRILLT 180
      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVDSATMANKGLIELIACILFNMEPD 240
      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVDSATMANKGLIELIACILFNMEPD 240
      241 QIEVVIHQSIHSMVDYVDSVLAQMGKPMRTPIAHAMAMPFDSGVAPLDIFEVGH 300
      241 QIEVVIHQSIHSMVDYVDSVLAQMGKPMRTPIAHAMAMPFDSGVAPLDIFEVGH 300
      301 MDEFKPDLRFPCLRLAYAIKSGGIMPTVLANANEIAVEAFINEEVFTDIAVIERSM 360
      301 MDEFKPDLRFPCLRLAYAIKSGGIMPTVLANANEIAVEAFINEEVFTDIAVIERSM 360

```

```

      301 MDEFKPDLRFPCLRLAYAIKSGGIMPTVLANANEIAVEAFINEEVFTDIAVIERSM 360
      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394
      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394

```

RESULT 3

```

US-09-941-947a-8
; Sequence 8, Application US/09941947a
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Bristowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Matchee
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Roviello, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947a
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Methylobionas 16a
US-09-941-947a-8

```

```

Query Match      100.0%; Score 1985; DB 11; Length 394;
Best Local Similarity 100.0%; Pred. No. 7,2e-186;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

      1 MKGICILGATGSIQVSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHREYAVVWMS 60
      1 MKGICILGATGSIQVSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHREYAVVWMS 60
      61 KVAEFKQRIASPVAADIKVLSGSEALQOVATLENDVTMAAIVGAAGLLPTLAAAKAGKT 120
      61 KVAEFKQRIASPVAADIKVLSGSEALQOVATLENDVTMAAIVGAAGLLPTLAAAKAGKT 120
      121 VILANKELVMSGOIFMOWSDSGAVLLPISSEHNAIQCPAGTTPGHTAKOARRILLT 180
      121 VILANKELVMSGOIFMOWSDSGAVLLPISSEHNAIQCPAGTTPGHTAKOARRILLT 180
      121 VILANKELVMSGOIFMOWSDSGAVLLPISSEHNAIQCPAGTTPGHTAKOARRILLT 180
      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVDSATMANKGLIELIACILFNMEPD 240
      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVDSATMANKGLIELIACILFNMEPD 240
      241 QIEVVIHQSIHSMVDYVDSVLAQMGKPMRTPIAHAMAMPFDSGVAPLDIFEVGH 300
      241 QIEVVIHQSIHSMVDYVDSVLAQMGKPMRTPIAHAMAMPFDSGVAPLDIFEVGH 300
      301 MDEFKPDLRFPCLRLAYAIKSGGIMPTVLANANEIAVEAFINEEVFTDIAVIERSM 360
      301 MDEFKPDLRFPCLRLAYAIKSGGIMPTVLANANEIAVEAFINEEVFTDIAVIERSM 360
      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394
      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394

```

RESULT 4

```

US-10-381-779-117
; Sequence 117, Application US/10381779

```

```

1 Publication No. US20030219798A1
2
3 GENERAL INFORMATION:
4
5 APPLICANT: Gokken, Ravi R
6
7 APPLICANT: Jessen, Holly
8
9 APPLICANT: Zidwick, Mary Jo
10
11 TITLE OF INVENTION: Isoprenoid Production
12
13 FILE REFERENCE: 12904/002051
14
15 CURRENT APPLICATION NUMBER: US/10/361,779
16
17 CURRENT FILING DATE: 2003-03-28
18
19 PRIOR APPLICATION NUMBER: PCT/US01/30338
20
21 PRIOR FILING DATE: 2001-09-28
22
23 PRIOR APPLICATION NUMBER: 60/236,580
24
25 PRIOR FILING DATE: 2000-09-29
26
27 NUMBER OF SEQ ID NOS: 190
28
29 SOFTWARE: PasteBQ for Windows Version 4.0
30
31 SEQ ID NO 117
32
33 LENGTH: 396
34
35 TYPE: PRT
36
37 ORGANISM: Pseudomonas aeruginosa
38
39 US-10-361-779-117

```

Query Match	61.2%;	Score 1215.5;	DB 12;	Length 396;
Best Local Similarity	64.1%;	Pred. No. 2.1e-110;		
Matches 245;	Conservative 48;	Mismatches 88;	Indels 1;	Gaps 1;

Qy	4	ICILGATGSI	GVSTLDDV	ARBPDKT	OVVALL	TANGNID	LYBQCLAH	DEYAVV	WVME	KVA	63
Dd	7	ISVLCNG	STIGSTLDDV	VRPDR	YEA	FALTY	FSFLAIE	ELCLHR	FVYAV	VEBQAAI	66
Qy	64	EFKOR	IAAPVAD	IKVLS	GBAL	QOAV	TLEN	DTVMA	IVGAG	LPLTAAKAK	123
Dd	67	ALOG	SIAAG	IR	TR	YFGR	QALCE	VASAP	EVDM	MAIVGAG	125
Qy	124	ANKEAL	VMSG	GF	FMQAV	SDSGAV	LLP	IDSE	NALIQ	CM	183
Dd	126	ANKEAL	VMSG	AL	FMQAV	KRS	GAV	LLP	IDSE	NALIQ	185
Qy	184	GPRPR	PIET	US	VP	DOA	VAP	KYMD	MG	KIS	243
Dd	186	GPRPR	ETPL	QGLAS	VTP	BOCA	HRP	NMS	WG	KIS	245
Qy	244	VVIH	POSI	HS	MWD	YD	VS	LAQ	MGN	PDK	303
Dd	246	VVIH	PO	SV	HS	MWD	YD	VS	LAQ	MGN	305
Qy	304	EK	DEL	K	F	P	C	L	R	A	363
Dd	306	Q	R	D	E	O	R	F	P	C	365
Qy	364	K	P	D	A	G	S	L	E	V	385
Dd	366	A	V	T	A	V	E	S	L	O	387

RESULT 5
 US-19-381-779-119
 Sequence 119, Application US/10381779
 Publication No. US20030219798A1
 GENERAL INFORMATION:
 APPLICANT: Gokarn, Ravi R
 APPLICANT: Jessen, Holly
 APPLICANT: Zidwick, Mary Jo
 TITLE OF INVENTION: Isoprenoid Production
 FILE REFERENCE: 129804/002US1
 CURRENT APPLICATION NUMBER: US/10/381,779
 CURRENT FILING DATE: 2003-03-28
 PRIOR APPLICATION NUMBER: PCT/US01/30328
 PRIOR FILING DATE: 2001-03-28
 PRIOR APPLICATION NUMBER: 60/236,580
 PRIOR FILING DATE: 2000-03-29
 NUMBER OF SEQ. ID NOS: 190
 SOFTWARE: PatsSeq for Windows Version 4.0
 SEQ ID NO 119

```

; LENGTH: 394
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-381-779-119

```

Query Match	53.0%;	Score 1052.5;	DB 12;	Length 394;
Best Local Similarity	55.5%;	Pred. No. 2,1e-94;		
Matches 216;	Conservative 49;	Mismatches 123;	Indels 1;	Gaps 1;

[illegible]

```

US-10-381-779-100
RESULT 6
? Sequence 100, Application US/10381779
? Publication No. US20030219798A1
? GENERAL INFORMATION:
? APPLICANT: GOKARN, Ravi R
? APPLICANT: Jessen, Holly
? APPLICANT: Zidwick, Mary Jo
? TITLE OF INVENTION: Isoprenoid Production
? FILE REFERENCE: 12904/002U51
? CURRENT APPLICATION NUMBER: US/10/381, 779
? CURRENT FILING DATE: 2003-03-28
? PRIOR APPLICATION NUMBER: PCT/US01/30328
? PRIOR FILING DATE: 2001-09-28
? PRIOR APPLICATION NUMBER: 60/736,580
? PRIOR FILING DATE: 2000-09-29
? NUMBER OF SEQ ID NOS: 190
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 100
? LENGTH: 398
? TYPE: PRT
? ORGANISM: Escherichia coli
US-10-381-779-100

```

	Query Match	52.6%;	Score 1045;	DB 12;	length 398;
	Best Local Similarity	54.9%;	Pred. No. 1.le-93;		
	Matches 219; Conservative	56;	Mismatches 118;	Indels 6;	Gaps 3
Qy	1 MKGICILIGATGSIGVSTLDVAHPDPKQVMALITANGIDALEYOCIAHHPAYAVVWS	60			
	: : : : : : : :				
Dd	1 MKQTLTSGTSIGCSITLADVARNPHERRVVALVGKNVTNRVEOCLFSPPRYAIVMDEBA	60			
Qy	61 KVAEKKRIITAAPVADIKVLSSSEALQQVATLENVDYMAALVGAAGLLPTLAAAKCKXT	120			
	: : : : : : : :				
Dd	61 S-AKLTKMTLQGGSGRTFVLTSCQAACDMALEDDVDYMAALVGNAGLLPTLAIRACKT	119			

```
QY 121 VLLANKKALVMSGQIFMVAQVSDGAVLLPIDSSEHNAIFQCM--PAGYTPGHTAKQAR--- 175
DB 120 ILLANKSESVTCRLPMDAVKOSKQALLPVDSSEHNAIFQSLPQPIQHNGYADLENGVAV 179
QY 176 RLLLTASGGFFRRPTLETLSVTPDQAVAHPKYDMGKRKISVDSATMMNKGLEIEACLPF 235
DB 180 SILLTSSGGFFRRPTLETLSVTPDQAVAHPKYDMGKRKISVDSATMMNKGLEIEACLPF 239
QY 236 NMEPDQIEVVIHPQSIHSHWVDYDGSVLAQMGNPDMKRTPIAHAMAMPERRFDSGVAPLDI 295
DB 240 NASASQMEVLIHPQSVIHSWVRXQDGSVLAQJGEPDMKRTPIAHAMAMPERRFDSGVAPLDI 299
QY 296 FEVGHMDFEKPDLKRPCLALAYEALIKSGGIMFTVANAANEIVAEFLNEBVKFTDIAYI 355
DB 300 CKLSALTFAAPDYDRPCLALAYEALIKSGGIMFTVANAANEIVAEFLAQOIRFTDIAYI 359
QY 356 IERSMAQFKPDAGSLVLYQADQDAFEVARDIITKTLVA 394
DB 360 NLSVLEKMDMREPOQCVDDVLSDVANNREVARKEVRLAS 398
```

```
RESULT 7
US-10-381-779-118
; Sequence 118, Application US/10381779
; Publication No. US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jessen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002US1
; CURRENT APPLICATION NUMBER: US/10/381,779
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 118
; LENGTH: 398
; TYPE: PRF
; ORGANISM: Escherichia coli
US-10-381-779-118
```

```
Query Match 52.6%; Score 1045; DB 12; Length 398;
Best Local Similarity 54.9%; Pred. No. 1.1e-93;
Matches 219; Conservative 56; Mismatches 118; Indels 6; Gaps 3;

QY 1 MKGICILGATGSIQVSTLDVVAHHPKRYQVVALTANGNIDALYEQCLAHHPREYAVVWMS 60
DB 1 MKQITLIGSTGSIQVSTLDVVAHHPKRYQVVALTANGNIDALYEQCLAHHPREYAVVWMS 60
QY 61 KVAEFKRIASPVADIKYLSGSEALQVATLENVDYMAAIVGAGLLPTLAAAKAGT 120
DB 61 S-AKLKTMLOQGSSTEVLSGQAAQADVALEDVQVAAIVGAGLLPTLAAAPAGKT 119
QY 121 VLLANKKALVMSGQIFMVAQVSDGAVLLPIDSSEHNAIFQCM--PAGYTPGHTAKQAR--- 175
DB 120 ILLANKSESVTCRLPMDAVKOSKQALLPVDSSEHNAIFQSLPQPIQHNGYADLENGVAV 179
QY 176 RLLLTASGGFFRRPTLETLSVTPDQAVAHPKYDMGKRKISVDSATMMNKGLEIEACLPF 235
DB 180 SILLTSSGGFFRRPTLETLSVTPDQAVAHPKYDMGKRKISVDSATMMNKGLEIEACLPF 239
QY 236 NMEPDQIEVVIHPQSIHSHWVDYDGSVLAQMGNPDMKRTPIAHAMAMPERRFDSGVAPLDI 295
DB 240 NASASQMEVLIHPQSVIHSWVRXQDGSVLAQJGEPDMKRTPIAHAMAMPERRFDSGVAPLDI 299
QY 296 FEVGHMDFEKPDLKRPCLALAYEALIKSGGIMFTVANAANEIVAEFLNEBVKFTDIAYI 355
DB 300 CKLSALTFAAPDYDRPCLALAYEALIKSGGIMFTVANAANEIVAEFLAQOIRFTDIAYI 359
```

```
QY 356 IERSMAQFKPDAGSLVLYQADQDAFEVARDIITKTLVA 394
DB 360 NLSVLEKMDMREPOQCVDDVLSDVANNREVARKEVRLAS 398
```

```
RESULT 8
US-10-381-779-122
; Sequence 122, Application US/10381779
; Publication No. US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jessen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002US1
; CURRENT APPLICATION NUMBER: US/10/381,779
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 122
; LENGTH: 405
; TYPE: PRF
; ORGANISM: Pasteurella multocida
US-10-381-779-122
```

```
Query Match 52.5%; Score 1042; DB 12; Length 405;
Best Local Similarity 53.2%; Pred. No. 2.3e-93;
Matches 210; Conservative 68; Mismatches 109; Indels 8; Gaps 3;
```

```
QY 1 MKGICILGATGSIQVSTLDVVAHHPKRYQVVALTANGNIDALYEQCLAHHPREYAVVWMS 60
DB 7 MKQITLIGSTGSIQVSTLDVVAHHPKRYQVVALTANGNIDALYEQCLAHHPREYAVVWMS 66
QY 61 KVAEFKRIASPVADIKYLSGSEALQVATLENVDYMAAIVGAGLLPTLAAAKAGT 120
DB 67 AAKMLAEKRAHQ-SQTVLAGQQAICELAHHPENDYMAAIVGAGLLPTLAAAKAGT 125
QY 121 VLLANKKALVMSGQIFMVAQVSDGAVLLPIDSSEHNAIFQCM--PAGYTPGHTAKQAR 174
DB 126 VLLANKKALVMSGQIFMVAQVSDGAVLLPIDSSEHNAIFQSLPQPIQHNGYADLENGVAV 184
QY 175 RLLLTASGGFFRRPTLETLSVTPDQAVAHPKYDMGKRKISVDSATMMNKGLEIEACLPF 234
DB 185 SKVLTSSGGFFRRPTLETLSVTPDQAVAHPKYDMGKRKISVDSATMMNKGLEIEACLPF 244
QY 235 NMEPDQIEVVIHPQSIHSHWVDYDGSVLAQMGNPDMKRTPIAHAMAMPERRFDSGVAPLDI 294
DB 245 FNASAEVLIHPQSVIHSWVRXQDGSVLAQJGEPDMKRTPIAHAMAMPERRFDSGVAPLDI 304
QY 295 IFVGHMDFEKPDLKRPCLALAYEALIKSGGIMFTVANAANEIVAEFLNEBVKFTDIAYI 354
DB 305 FYQLNGLFIEBPDYRYPCCLALADAFSAGYATTMAANEIVAEFLNEBVKFTDIAYI 364
QY 355 IERSMAQFKPDAGSLVLYQADQDAFEVARDIITKTLVA 389
DB 365 LNLQVVSILQPOKIHCHIEDVLEVKRAEELSII 399

RESULT 9
US-10-381-779-99
; Sequence 99, Application US/10381779
; Publication No. US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jessen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002US1
```

```

1 CURRENT APPLICATION NUMBER: 2003-03-28
2 CURRENT FILING DATE: 2003-09-28
3 PRIOR APPLICATION NUMBER: PCT/US01/30328
4 PRIOR FILING DATE: 2001-09-28
5 PRIOR APPLICATION NUMBER: 60/236,580
6 PRIOR FILING DATE: 2000-09-29
7 NUMBER OF SEQ ID NOS: 190
8 SOFTWARE: FastSeq for Windows Version 4.0
9 SEQ ID NO: 99
10 LENGTH: 397
11 TYPE: PRN
12 ORGANISM: Haemophilus influenzae
13
14 US-10-361-779-99

```

Query Match	50.7%;	Score 1007;	DB 12;	Length 397;
Best Local Similarity	50.6%;	Pred. No. 6e-90;		
Matches 200;	Conservative 76;	Mismatches 111;	Indels 8;	Gaps 3;

QY	KGI C I G A G Y S I E V N T L D V A A H P R D U Q V A L T A N G I N D A L Y E O G I L A H N F E A V V W E S K	61
Db	4 Q N I V I L A S G S I E K S T L S Y I E N N P O K T A F A L V G K N V E A N F B O C I F R P R F A L D V N A	63
QY	V A E F K O R I A S P A D I K V L S G S E A L Q O V A T L E N V T M A A I V G A G L I P T L A A K A K T V	121
Db	64 A K L R E K L A H N I - P R E V L A G R R A I C E L A A H P D A Q I M A S I V G A A G I L I T I S A Y A K R V	122
QY	122 L I A N K A L V A S G O I E M O A V S D S G A V L I P D S E H N A I P O C M P -----A G Y T P G T A N A O A R	175
Db	123 L I A N K S I L V P C G L P I D A V A N Y G S K L L P Y D S E H N A I P O S L P P E A O E K I G F C P -L S E I G V S	181
QY	176 R I L L T S G G F R T P T J E L I S S V T P Q O A V A N H P X M D G R K I S V D A S T M N M K G L E L E A C L F	235
Db	182 K I I L T S G G F R T P T J E L I S S V T P Q O A V A N H P M S M G K I I S V D A S T M N M K G L E L E A C L F	241
QY	N M E P D O L E V I T I H S M V D Y D C S V L A O K M P D M T P L A H A M A W E R P D S G A V L D I	295
Db	242 N A S A E M E V I I H Q S I I H S M V R Y D G S V I T T Q G N D M E T P L A E T A Y A H R F R A G V E L D F	301
QY	296 F E V G H M D F E R E D L K R P C L R L A Y E A I K S G I A P T V L A N A I E A V A F T N E E K T D I A V I	355
Db	302 F K I E L L F I E P D N R Y P N I K L A I D A P A Q V A T T M A N A N E I A V A Q A F L D R Q I G M D I A K I	361
QY	356 I E R S M A F K R P D G S L E L V L Q A D Q O D A R E V A R I O I K	390
Db	362 N S K T I E R I S E Y T I Q N I D D V L E I D A Q R E L I A K T I Q R	396

```

US-10-381-779-120
RESULT 10
Sequence 120. Application US/10381779
Publication NO. US20030219798A1
GENERAL INFORMATION:
APPLICANT: Gokarrn, Ravi R
APPLICANT: Jessen, Holly Jo
APPLICANT: Zidwick, Mary Jo
TITLE OF INVENTION: Isoprenoid Production
FILE REFERENCE: 12904/002US1
CURRENT APPLICATION NUMBER: US/10/381, 779
CURRENT FILING DATE: 2003-03-28
PRIOR APPLICATION NUMBER: PCT/US01/30328
PRIOR FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: 60/236, 580
PRIOR FILING DATE: 2000-09-29
NUMBER OF SEQ ID NOS: 190
SOFTWARE: FastSeq for Windows Version 4.0.0
SEQ ID NO 120
LENGTH: 397
TYPE: PKT
ORGANISM: Haemophilus influenzae
US-10-381-779-120

```

Query Match	Score 1007;	DB 12;	Length 397
Best Local Similarity	50.7%;	Pred. No. 6e-90;	

Matches 200; Conservative 76; Mismatches 111; Indels 8; Gaps 3;

[illegible]

```

RESULT 11
US-10-381-779-101
; Sequence 101, Application US/10381779
; Publication No. US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jesseen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002US1
; CURRENT APPLICATION NUMBER: US/10/381,779
; CURRENT FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 101
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Zymonas mobilis
US-10-381-779-101

```

	Query Match	50.4%; Score 999.5; DB 12; Length 388;
	Best Local Similarity	Pred. No. 3.2e-89;
	Matches 199; Conservative 74; Mismatches 109; Indels 9; Gaps 3	
QY	4 ICILGANGSGTGVSTLDVVARHPDKXYOVATLNGCIDLYQCCLAHHEPAVVVMESKVA	63
DB	7 VTVVGAGISGHSTDLIERNDPRYOVTALNNRNVKOLADAKRTNKGAVINADPSLYN	66
QY	64 EFQRIRASVPADIKYLSCSEALQQVATLENDVTMAAIVGAGLLPTLAATAKAGTVLL	122
DB	67 DLKEALLGSG---SVESAAGADALVLRAMN--GADWTMAAIICAGIKATLTAAIRGKTVAL	122
QY	124 ANKEALVMSGOIFPMOAVSDGVALPLPISEHNALIPOCAPAYTGHTKQARRILLTSAG	183
DB	123 ANKESLVASAGGLMIDAVRHGTLLLPPVSERNHALPQCF-----PHHNRDVRRIITITSAG	177
QY	184 GPFRPTIETLSVTPDQAVANPKMDNRKRISVDSATNMKGELIEACLLFNNEPDQIE	244

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 11.5928 Seconds
(without alignments)
3268.453 Million cell updates/sec

Title: US-09-941-947a-8
Perfect score: 1985
Sequence: 1 MKGICILGATGSGIGVSTLDV.....LQADQAREVARDIKITLVA 394

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: pirl:*
2: pirl:*
3: pirl:*
4: pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1215.5	61.2	396	2 E83188	1-deoxy-D-xyloose
2	1052.5	53.0	394	2 B82000	1-deoxy-D-xyloose
3	1045	52.6	398	2 B64741	YaeM protein - Bac
4	1044	52.6	398	2 G90650	1-deoxy-D-xyloose
5	1044	52.6	398	2 G85501	1-deoxy-D-xyloose
6	1040.5	52.4	394	2 A81229	1-deoxy-D-xyloose
7	1027	51.7	398	2 A81229	1-deoxy-D-xyloose
8	1023	51.5	402	2 D82099	1-deoxy-D-xyloose
9	1020	51.4	398	2 AG0128	1-deoxy-D-xyloose
10	1007	50.7	397	2 A64014	conserved hypotet
11	991	47.4	396	2 H87228	1-deoxy-D-xyloose
12	892.5	45.0	394	2 S76331	hypothetical prote
13	891	44.9	398	2 F84957	1-deoxy-D-xyloose
14	865	43.6	399	2 AH2349	deoxyxyloose 5-ph
15	857.5	43.2	399	2 H87486	1-deoxy-D-xyloose
16	854	43.0	395	2 E97121	1-deoxy-D-xyloose
17	849	42.8	397	2 AC2897	1-deoxy-D-xyloose
18	849	42.8	397	2 E97572	1-deoxy-D-xyloose
19	818.5	41.2	406	2 T52570	1-deoxy-d-xyloose
20	816.5	41.1	380	2 AE1239	deoxyxyloose 5-ph
21	812.5	40.9	380	2 A11601	1-deoxy-D-xyloose
22	800	40.3	365	2 B83522	1-deoxy-d-xyloose
23	798.5	40.2	388	2 B69881	conserved hypotet
24	709.5	35.7	379	2 B81578	1-deoxy-D-xyloose
25	707.5	35.6	379	2 B72091	CT071 hypothetical
26	706.5	35.6	379	2 G86533	1-deoxy-D-xyloose
27	698	35.2	396	2 F75388	hypothetical prote
28	691	34.8	396	2 A70923	1-deoxy-D-xyloose
29	684.5	34.5	379	2 G81712	1-deoxy-D-xyloose

681	34.3	406	2	A87107	hypothetical prote
646.5	32.6	379	2	A71562	hypothetical prote
641	32.3	380	2	F70336	conserved hypotet
629	31.7	383	2	C71304	conserved hypotet
547	27.6	376	2	B72321	conserved hypotet
486	24.5	356	2	B81278	probable 1-deoxy-D
473	23.8	368	2	H64546	conserved hypotet
455	22.9	368	2	G71961	hypothetical prote
443	22.7	443	2	S03257	methyl coenzyme M
443	22.7	443	2	S03257	probable gamma-glu
443	22.7	443	2	S03257	fatty-acid synthas
443	22.7	443	2	S03257	prephenate dehydro
443	22.7	443	2	S03257	conserved hypotet
443	22.7	443	2	S03257	hypothetical prote
443	22.7	443	2	S03257	periodic cryptopha
443	22.7	443	2	S03257	prephenate dehydro

ALIGNMENTS

RESU	1	681	34.3	406	2	A87107	hypothetical prote
E831	1-dec	Y-d-xyloose 5-phosphate reductoisomerase PA3650 [imported] - Pseudomonas aeruginos					
C:Spa	ies: Pseudomonas aeruginosa						
C:Dat	: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000						
C:Acc	sion: E83188						
R:Stc	er, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Br						
Adman	S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim,						
adman	Y., S., Olson, M.V.						
Natun	406, 959-964, 2000						
A:Tit	e: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho						
A:Ref	rence number: A82950; PMID:20437337; PMID:10984043						
A:Acc	sion: E83188						
A:Stc	us: preliminary						
A:Mo	cule type: DNA						
A:Ref	dues: 1-396 <STO>						
A:Cre	s-references: GB:AE004785; GB:AE004091; NID:G9949809; PIDN:AA07038.1; GSPDB:GN001						
A:Ex	perimental source: strain PA01						
C:Gen	tics:						
A:Gen	: dxt; PA3650						
C:Seq	family: conserved hypothetical protein H1807						
Que	Y Match	61.2%	Score 1215.5;	DB 2;	Length 396;		
Be:	Local Similarity	64.1%;	Pred. No. 4.9e-84;				
Ma:	hes 245;	Conservative 48;	Mismatches 88;	Indels 1;	Gaps 1;		
QY	4	ICILANTSGIGVSTLDVYABHPDKQVVALFANGNIDALYBOCLAHHPYAVVMESKVA	63				
DB	7	ISVLANTSGIGVSTLDVYABHPDKQVVALFANGNIDALYBOCLAHHPYAVVMESKVA	66				
QY	64	EKKRIASPVADIVYLSGSEALQCVATLENDVTYMAAIVGAAGLLPTLAARAKGTVLL	123				
DB	67	ALQGSILAAAGIR-TRVLFEQALCEVASAPEYDYMMAAIVGAAGLPSTLAAEAGRVLL	125				
QY	124	AKKEALVMSGQIFPMQAVDSDGAVLLPISSEHNAIFQCPAGTTPGHTLAKQARRILLTSAG	183				
DB	126	AKKEALVMSGALFPMQAVRSQAVLLPISSEHNAIFQCPAGTTPGHTLAKQARRILLTSAG	185				
QY	184	GPFRPTPIETLSVTPDQAVAHPKVMDGKRIKISVDSATVMNKGLELIEACLPFMNEPDQTE	243				
DB	186	GPFRPTPIETLSVTPDQAVAHPKVMDGKRIKISVDSATVMNKGLELIEACLPFMNEPDQTE	245				
QY	244	VTHIQSIIHSMVDVVDGSLAOKNGNPMKPTIAPAMMPERFDSGVAPLDIFEEGHMF	303				
DB	246	VTHIQSIIHSMVDVVDGSLAOKNGNPMKPTIAPAMMPERFDSGVAPLDIFEEGHMF	305				
QY	304	EKPDKRFPCLLAEYAIKSGIMPTVLAANEIAVEAFINEEVYPTDIIVTIERSMAQF	363				
DB	306	QKPDQRFPCLLAEYAIKSGIMPTVLAANEIAVEAFINEEVYPTDIIVTIERSMAQF	365				
QY	364	KPDQAGSELEVTLQADQAREVA 385					

DB 366 AVTAVESLDQVLAADRARSVA 367

RESULT 2

B82000
1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.-) NMA0083 [imported] - Neisseria meningitidis

C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: B82000

R:Parthill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

R:Holroyd, J.; Jagsall, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A:Reference number: A81775; M01D:2022556; PMID:10761919

A:Accession: B82000

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <PAR>

A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83399.1; PID:g737885

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: dxt; NMA0083

C:Superfamily: conserved hypothetical protein HI0807

C:Keywords: oxidoreductase

Query Match 53.0%; Score 1052.5; DB 2; Length 394;
Best Local Similarity 55.5%; Pred. No. 9.5e-72;
Matches 216; Conservative 49; Mismatches 123; Indels 1; Gaps 1;

4 ICILGATSGISGVSTLDVVARHPDKYQVVALFRANGNIDALVEQCLAHPEYAVVNESKVA 63

6 LTLISSTGSIIGSTLDVVARHPDKYQVVALFRANGNIDALVEQCLAHPEYAVVNESKVA 65

64 EKQRIASPVADIKYLSSEALQVATLNDVTMAAIVAGAILPTLAARKAKTYL 123

66 RLEALKRDGTA-TYVLAHQALVDAASDEVSGWCAVAGVLPALAAQKRTYL 124

124 AKKALVMSGQIFMOWVDSGAVLLPIDEENAFQCPDAGTTPGHTAKQARRILLTASG 183

125 AKKALVMSGQIFMOWVDSGAVLLPIDEENAFQCPDAGTTPGHTAKQARRILLTASG 184

184 GFPRTPPIETLSVTPDQAVARHPDKYQVVALFRANGNIDALVEQCLAHPEYAVVNESKVA 243

185 GFPRTPPIETLSVTPDQAVARHPDKYQVVALFRANGNIDALVEQCLAHPEYAVVNESKVA 244

244 VVIHPSIISHWVDYDGSVLAQMGNDKRTPIAHAMAPERFDSVAPLDFEYGHDF 303

245 VVIHPSIISHWVDYDGSVLAQMGNDKRTPIAHAMAPERFDSVAPLDFEYGHDF 304

304 EKEDLKRFPCLRLAYEAIKSGIMPTLVANAEIYAEAFINEEVKFTDIAVIERSMQF 363

305 QKPDFRFPCLRLAYEAIKSGIMPTLVANAEIYAEAFINEEVKFTDIAVIERSMQF 364

364 KPDGASLELVQAQODARVARDIKTL 392

365 FSDGIGDIGGLLAQDARTAAQAPFGLT 393

RESULT 3

B64741

YaeM protein - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: B64741

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; M01D:97426617; PMID:9278503

A:Accession: B64741

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-398 <BLAT>

A:Cross-references: GB:AE000126; GB:U00096; NID:g1786358; PIDN:AACT3284.1; PID:g1786369;

A:Experimental source: strain K-12, substrain MGI655

C:Genetics:

A:Gene: yaeM

C:Superfamily: conserved hypothetical protein HI0807

C:Keywords: nucleotide binding; P-loop

F11-119/Region: nucleotide-binding motif A (P-loop)

Query Match 52.6%; Score 1045; DB 2; Length 398;

Best Local Similarity 54.9%; Pred. No. 3.5e-71;

Matches 219; Conservative 56; Mismatches 118; Indels 6; Gaps 3;

1 MKGICILGATSGISGVSTLDVVARHPDKYQVVALFRANGNIDALVEQCLAHPEYAVVNES 60

1 MKGILITGSGISGVSTLDVVARHPDKYQVVALFRANGNIDALVEQCLAHPEYAVVNES 60

61 KVAEFKRIASPVADIKYLSSEALQVATLNDVTMAAIVAGAILPTLAARKAKGT 120

61 S-AKLITKTMLOQGSQSRREVLSGQACDMALBDVDQVMAIVAGAILPTLAIRAGKT 119

121 VILANKALVMSGQIFMOWVDSGAVLLPIDEENAFQCPDAGTTPGHTAKQARRILLTASG 175

120 ILLANKSLVTCGLFMDAVKQSKQLLPVDSERNALFQSLPQTLQNLGPADEQNGV 179

176 RIILTAGCGPRFRTPPIETLSVTPDQAVARHPDKYQVVALFRANGNIDALVEQCLAHPEYAVVNES 235

180 SILITGSGPRFRTPPIETLSVTPDQAVARHPDKYQVVALFRANGNIDALVEQCLAHPEYAVVNES 239

236 MNEPDIEVTHPOSTIHSWVDYDGSVLAQMGNDKRTPIAHAMAPERFDSVAPLDFEYGHDF 295

240 NASASQMEVLIHPQSVIHSWVRVODGSVLAQMGNDKRTPIAHAMAPERFDSVAPLDFEYGHDF 299

296 FEVGHMDFKRPCLRLAYEAIKSGIMPTLVANAEIYAEAFINEEVKFTDIAVIERSMQF 355

300 CKSLATFPADYRYCYCLKLMARFPGQALTAANAETITAAIQAQIRTDIAL 359

356 IERSMAQKPDGASLELVQAQODARVARDIKTLVA 394

360 NLVLEKMDKMEPQCDVDVLSVDANAREVARKVEVRLAS 398

RES T 4
G300
1-d xy-D-xylulose 5-phosphate reductoisomerase [imported] - Escherichia coli (strain O1

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C:Accession: G90650

R:Harash, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gskara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

es. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; M01D:21156231; PMID:11258796

A:Accession: G90650

A:Molecule type: DNA

A:Status: preliminary

A:Cross-references: GB:BA000007; PIDN:BA833598.1; PID:g13359631; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain R1MD 0505952

C:Genetics:

A:Gene: ECs0175

C:Superfamily: conserved hypothetical protein HI0807

Query Match 52.6%; Score 1044; DB 2; Length 398;

Best Local Similarity 54.9%; Pred. No. 4.2e-71; Indels 6; Gaps 3;

Matches 219; Conservative 55; Mismatches 119;

1 MKGICILGATSGISGVSTLDVVARHPDKYQVVALFRANGNIDALVEQCLAHPEYAVVNES 60

1 MKGILITGSGISGVSTLDVVARHPDKYQVVALFRANGNIDALVEQCLAHPEYAVVNES 60

61 KVAEFKRIASPVADIKYLSSEALQVATLNDVTMAAIVAGAILPTLAARKAKGT 120

61 S-AKLITKTMLOQGSQSRREVLSGQACDMALBDVDQVMAIVAGAILPTLAIRAGKT 119

QY 121 VLLANKKALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 175
 DB 120 ILLANKKESLVTCGRFLPMDAVKOSKQOLLFPVDSHNAIFQSLPQPIQHNLYGADLEQNGV 179
 QY 176 RLLTASGGPFRRPTLETLSVTPDOAVAHFPMKDMGRKISVDSATMMNKGLEIELACLLF 235
 DB 180 SILLTSGSGPFRRPTLETLSVTPDOACHPMKDMGRKISVDSATMMNKGLEIELARMLF 239
 QY 236 NMEPDQIEVVIHPQSIHSMVDYVDSVLAQKGNPDNRPPIAHAMAMPERRDSGVAPLDI 295
 DB 240 NASAQMEVLHPQSVIHSMTVYQDSVLAQGEEDMRTPIAHAMAMPERRVNSGVKPLDF 299
 QY 296 FEVGHMDFEKPDLKRPFCRLAYEAIKSGIIPVTYLANANEIAYEAFLEBEVKFTDIYVI 355
 DB 300 CKLSALTFAPDYDRPCIKLAMEAFEGQAATTLNANETLVAALFQAQIRFTDIAL 359
 QY 356 IERSNAQFPDAGSLBLVLAQDQAREVARDIITLVA 394
 DB 360 NLVLEKMKDMRERPCQVDVLSVDASAREVARKVRLAS 398

RESULT 5

G85501
 1-deoxy-D-xylulose 5-phosphate reductoisomerase [similarity] - Escherichia coli (strain
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Nov-2001
 C:Accession: G85501
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Gotohbeck, E.J.; Davis, N.W.; Lin, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: AB5480; MUID:21074935; PMID:11206551
 A:Accession: G85501
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-398 <STO>
 A:Cross-references: GB:AB005174; NID:912512898; PIDN:AMG54475.1; GSPDB:GN00145; UMGF:Z01
 A:Experimental source: strain O157:H7, substrain ED1933
 C:Genetics:
 A:Gene: yaeM
 C:Superfamily: conserved hypothetical protein HI0807

Query Match 52.6%; Score 1044; DB 2; Length 398;
 Best local Similarity 54.9%; Pred. No. 4, 2e-71;
 Matches 219; Conservative 55; Mismatches 119; Indels 6; Gaps 3;
 QY 1 MGICIGATGSGITLDDVARRHPDKYOVVALTANGNIDALYEQCLAHPEYAVVMS 60
 DB 1 MGOITLISGSGITLDDVARRHPDKYOVVALTANGNIDALYEQCLAHPEYAVVMS 60
 QY 61 KVAEFKRIASPVADIKVLGSGEALQOVATLENDVYMAIVGAAGLLPTLAAKAGKT 120
 DB 61 S-AKLKLTWLQOOGSRTEVLSCGOAAADMALEEDVQWMAIVGAAGLLPTLAAIRAKT 119
 QY 121 VLLANKKALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 175
 DB 120 ILLANKKESLVTCGRFLPMDAVKOSKQOLLFPVDSHNAIFQSLPQPIQHNLYGADLEQNGV 179
 QY 176 RLLTASGGPFRRPTLETLSVTPDOAVAHFPMKDMGRKISVDSATMMNKGLEIELACLLF 235
 DB 180 SILLTSGSGPFRRPTLETLSVTPDOACHPMKDMGRKISVDSATMMNKGLEIELARMLF 239
 QY 236 NMEPDQIEVVIHPQSIHSMVDYVDSVLAQKGNPDNRPPIAHAMAMPERRDSGVAPLDI 295
 DB 240 NASAQMEVLHPQSVIHSMTVYQDSVLAQGEEDMRTPIAHAMAMPERRVNSGVKPLDF 299
 QY 296 FEVGHMDFEKPDLKRPFCRLAYEAIKSGIIPVTYLANANEIAYEAFLEBEVKFTDIYVI 355
 DB 300 CKLSALTFAPDYDRPCIKLAMEAFEGQAATTLNANETLVAALFQAQIRFTDIAL 359
 QY 356 IERSNAQFPDAGSLBLVLAQDQAREVARDIITLVA 394

DB 360 NLVLEKMKDMRERPCQVDVLSVDASAREVARKVRLAS 398

RESULT 6

AB1229
 1-deoxy-D-xylulose 5-phosphate reductoisomerase NMB0184 [imported] - Neisseria meningitidis
 C:Species: Neisseria meningitidis
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: AB1229
 R:Kretzlin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Bisen, J.A.
 Hickey, E.K.; Hatt, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 Krieger, H.; Yamashiro, Y.; Gill, J.; Scarlato, V.; Maignani, V.; Piazza, M.
 Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappaport, R.; Ver
 A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: AB1000; MUID:20175755; PMID:10710307
 A:Accession: AB1229
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-394 <TEF>
 A:Cross-references: GB:AE002375; GB:AE002098; NID:97225394; PIDN:AAE40641.1; PID:9722540;
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB0184
 C:Superfamily: conserved hypothetical protein HI0807

Query Match 52.4%; Score 1040.5; DB 2; Length 394;
 Best local Similarity 55.3%; Pred. No. 7, 6e-71;
 Matches 215; Conservative 47; Mismatches 126; Indels 1; Gaps 1;
 QY 4 ICIGATGSGITLDDVARRHPDKYOVVALTANGNIDALYEQCLAHPEYAVVMS 63
 DB 6 LTIISGSGITLDDVARRHPDKYOVVALTANGNIDALYEQCLAHPEYAVVMS 65
 QY 64 EFKRIASPVADIKVLGSGEALQOVATLENDVYMAIVGAAGLLPTLAAKAGKT 123
 DB 64 RLAKLKRQDGTATVLAHGAQALVDVADSGVMAIVGAAGLLPTLAAKAGKT 124
 QY 124 ANKALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCMAPGYTPGHTAKQARILLTASG 183
 DB 125 ANKELTVMSGALFMEETARANGAAVLVPDSEHNAIFQVLPDYAGRLNEHGIASILLTASG 184
 QY 184 GPERPTLETLSVTPDOAVAHFPMKDMGRKISVDSATMMNKGLEIELACLLFENNEPOIE 243
 DB 185 GPFLLADLAFTRITPAQAVKHPMWRMGKISVDSATMMNKGLEIELAHWLFNCPDRL 244
 QY 244 VVIHPQSIHSMVDYVDSVLAQKGNPDNRPPIAHAMAMPERRDSGVAPLDI 303
 DB 245 VVIHPQSVIHSMTVYQDSVLAQGEEDMRTPIAHAMAMPERRVNSGVKPLDF 304
 QY 304 EKPDLKRPFCRLAYEAIKSGIIPVTYLANANEIAYEAFLEBEVKFTDIYVI 363
 DB 305 QKPPDRFPFCRLAYEAIKSGIIPVTYLANANEIAYEAFLEBEVKFTDIYVI 364
 QY 364 KPDDAGSLBLVLAQDQAREVARDIITLVA 392
 DB 365 FSDGIGIGLADQARTRAARAFITL 393

RESULT 7

AF052
 1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.-) [imported] - Salmonella enterica
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AF0529
 R:Parfitt, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
 Threlkeld, S.; Coulter, S.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 Nature 413, 848-852, 2001
 A:Authors: Parfitt, J.; Coulter, S.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
 A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AF0529
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-398 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CMD08678.1; PID:g16501501; GSPDB:GN00176
 C:Genetics:
 A:Gene: dxt
 C:Superfamily: conserved hypothetical protein HI0807
 C:Keywords: oxidoreductase

Query Match 51.7%; Score 1027; DB 2; Length 398;
 Best Local Similarity 55.8%; Pred. No. 8e-70;
 Matches 223; Conservative 49; Mismatches 104; Indels 24; Gaps 7;

QY 1 MKGICILGATGSGTLDVVAHDPKRYQVVALTANGNIDALYECCLAHHPYAVVMS 60
 DB 1 MKGICILGATGSGTLDVVAHDPKRYQVVALTANGNIDALYECCLAHHPYAVVMS 60
 QY 61 KVAEFKQRIASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAARAGKT 120
 DB 61 KVAEFKQRIASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAARAGKT 120
 QY 61 SAQKIMQOHG-SKTEVLSSQACERAAALDEGVHVAIVGAGLPTLAARAGKT 119
 DB 61 SAQKIMQOHG-SKTEVLSSQACERAAALDEGVHVAIVGAGLPTLAARAGKT 119
 QY 121 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPM-----AGYTP---GHT 170
 DB 121 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPM-----AGYTP---GHT 170
 QY 120 ILANKESILVTGRLFMDEVKRSNARLLPIDSEHNAIFQSLPQIGHNIGYADLEONGVT 179
 DB 120 ILANKESILVTGRLFMDEVKRSNARLLPIDSEHNAIFQSLPQIGHNIGYADLEONGVT 179
 QY 171 AKGARILLTAGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 230
 DB 171 AKGARILLTAGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 230
 QY 180 S-----ILLTSGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 234
 DB 180 S-----ILLTSGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 234
 QY 231 ACULFNMEDPQIEVVIHPOSIHSMVDYVDGSLAQMGKNDKRTPIAHAMAFPERFDSGV 290
 DB 231 ACULFNMEDPQIEVVIHPOSIHSMVDYVDGSLAQMGKNDKRTPIAHAMAFPERFDSGV 290
 QY 235 ARVLFNASARQMEVLIHPOSVIHSVRKYGQSVLAQLGEPDKRTPIAHAMAFPERFDSGA 294
 DB 235 ARVLFNASARQMEVLIHPOSVIHSVRKYGQSVLAQLGEPDKRTPIAHAMAFPERFDSGA 294
 QY 291 APADIFEVGHMDPEKDLRFPCRLRAYEAIKSGGIMPTVLAANAIYEAFLNEBVKFT 350
 DB 291 APADIFEVGHMDPEKDLRFPCRLRAYEAIKSGGIMPTVLAANAIYEAFLNEBVKFT 350
 QY 295 QPDLFCULSLATYSADYQRYPCRLKAMEAFEGQAATTAANAIYEAFLAQCIKFT 354
 DB 295 QPDLFCULSLATYSADYQRYPCRLKAMEAFEGQAATTAANAIYEAFLAQCIKFT 354
 QY 351 DIA-----VILERSMAQFKPDAGSLVLAQADQADAREVAR 386
 DB 351 DIA-----VILERSMAQFKPDAGSLVLAQADQADAREVAR 386
 QY 355 DINGMLAVLER-NDLHP---ASVDVLAQVDAIAREVAR 390
 DB 355 DINGMLAVLER-NDLHP---ASVDVLAQVDAIAREVAR 390

RESULT 8

1-Deoxy-D-xylofucose 5-phosphate reductoisomerase VC2254 [imported] - Vibrio cholerae (str

C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: D82099
 C:Residues: 1-402 <HEI>

A:Cross-references: GB:AE004297; GB:AE003852; MID:g9656810; PIDN:AAFP95398.1; GSPDB:GN001
 A:Experimental source: serogroup O1, strain N16961; biotype E1 Tor
 C:Genetics:
 A:Gene: VC2254
 A:Map position: 1
 C:Superfamily: conserved hypothetical protein HI0807

Query Match 51.5%; Score 1023; DB 2; Length 402;
 Best Local Similarity 52.7%; Pred. No. 1.6e-69;
 Matches 214; Conservative 65; Mismatches 103; Indels 24; Gaps 4;

QY 1 MKGICILGATGSGTLDVVAHDPKRYQVVALTANGNIDALYECCLAHHPYAVVMS 60
 DB 1 MKGICILGATGSGTLDVVAHDPKRYQVVALTANGNIDALYECCLAHHPYAVVMS 60
 QY 61 KVAEFKQRIASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAARAGKT 120
 DB 61 KVAEFKQRIASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAARAGKT 120
 QY 61 SAQKIMQOHG-SKTEVLSSQACERAAALDEGVHVAIVGAGLPTLAARAGKT 119
 DB 61 SAQKIMQOHG-SKTEVLSSQACERAAALDEGVHVAIVGAGLPTLAARAGKT 119
 QY 121 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPM-----AGYTP---GHT 170
 DB 121 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPM-----AGYTP---GHT 170
 QY 120 ILANKESILVTGRLFMDEVKRSNARLLPIDSEHNAIFQSLPQIGHNIGYADLEONGVT 179
 DB 120 ILANKESILVTGRLFMDEVKRSNARLLPIDSEHNAIFQSLPQIGHNIGYADLEONGVT 179
 QY 171 AKGARILLTAGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 230
 DB 171 AKGARILLTAGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 230
 QY 180 S-----ILLTSGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 234
 DB 180 S-----ILLTSGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 234
 QY 231 ACULFNMEDPQIEVVIHPOSIHSMVDYVDGSLAQMGKNDKRTPIAHAMAFPERFDSGV 290
 DB 231 ACULFNMEDPQIEVVIHPOSIHSMVDYVDGSLAQMGKNDKRTPIAHAMAFPERFDSGV 290
 QY 235 ARVLFNASARQMEVLIHPOSVIHSVRKYGQSVLAQLGEPDKRTPIAHAMAFPERFDSGA 294
 DB 235 ARVLFNASARQMEVLIHPOSVIHSVRKYGQSVLAQLGEPDKRTPIAHAMAFPERFDSGA 294
 QY 291 APADIFEVGHMDPEKDLRFPCRLRAYEAIKSGGIMPTVLAANAIYEAFLNEBVKFT 350
 DB 291 APADIFEVGHMDPEKDLRFPCRLRAYEAIKSGGIMPTVLAANAIYEAFLNEBVKFT 350
 QY 295 QPDLFCULSLATYSADYQRYPCRLKAMEAFEGQAATTAANAIYEAFLAQCIKFT 354
 DB 295 QPDLFCULSLATYSADYQRYPCRLKAMEAFEGQAATTAANAIYEAFLAQCIKFT 354
 QY 351 DIA-----VILERSMAQFKPDAGSLVLAQADQADAREVAR 386
 DB 351 DIA-----VILERSMAQFKPDAGSLVLAQADQADAREVAR 386
 QY 355 DINGMLAVLER-NDLHP---ASVDVLAQVDAIAREVAR 390
 DB 355 DINGMLAVLER-NDLHP---ASVDVLAQVDAIAREVAR 390

QY 61 KVAEFKQRIASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAARAGKT 120
 DB 61 KVAEFKQRIASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAARAGKT 120
 QY 61 SAQKIMQOHG-SKTEVLSSQACERAAALDEGVHVAIVGAGLPTLAARAGKT 119
 DB 61 SAQKIMQOHG-SKTEVLSSQACERAAALDEGVHVAIVGAGLPTLAARAGKT 119
 QY 121 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPMAGYTPGHTAKOAR----- 175
 DB 121 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPMAGYTPGHTAKOAR----- 175
 QY 120 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPMAGYTPGHTAKOAR----- 172
 DB 120 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPMAGYTPGHTAKOAR----- 172
 QY 176 -----RILITAGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 228
 DB 176 -----RILITAGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 228
 QY 173 LSGHIDHILLTSGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 232
 DB 173 LSGHIDHILLTSGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 232
 QY 229 IEACILFNMEDPQIEVVIHPOSIHSMVDYVDGSLAQMGKNDKRTPIAHAMAFPERFDS 288
 DB 229 IEACILFNMEDPQIEVVIHPOSIHSMVDYVDGSLAQMGKNDKRTPIAHAMAFPERFDS 288
 QY 233 IEAKLFTSREQLKVLHPOSVIHSVRKYGQSVLAQLGEPDKRTPIAHAMAFPERFDS 292
 DB 233 IEAKLFTSREQLKVLHPOSVIHSVRKYGQSVLAQLGEPDKRTPIAHAMAFPERFDS 292
 QY 289 GVAPLDFEVGHMDPEKDLRFPCRLRAYEAIKSGGIMPTVLAANAIYEAFLNEBVK 348
 DB 289 GVAPLDFEVGHMDPEKDLRFPCRLRAYEAIKSGGIMPTVLAANAIYEAFLNEBVK 348
 QY 293 GVPALDFEVGHMDPEKDLRFPCRLRAYEAIKSGGIMPTVLAANAIYEAFLNEBVK 352
 DB 293 GVPALDFEVGHMDPEKDLRFPCRLRAYEAIKSGGIMPTVLAANAIYEAFLNEBVK 352
 QY 349 FTDLAVIERSMAQFKPD-----AGSLVLAQADQADAREVAR 390
 DB 349 FTDLAVIERSMAQFKPD-----AGSLVLAQADQADAREVAR 390
 QY 353 FTDLAVIERSMAQFKPD-----AGSLVLAQADQADAREVAR 390
 DB 353 FTDLAVIERSMAQFKPD-----AGSLVLAQADQADAREVAR 390

RESULT 9

1-Deoxy-D-xylofucose 5-phosphate reductoisomerase (EC 1.1.1.-) [imported] - Yersinia pestis

C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001

C:Accession: AG0128
 C:Residues: 1-398 <KIR>

A:Cross-references: GB:AL590842; PIDN:CAG89890.1; PID:g15979115; GSPDB:GN00175
 A:Experimental source: strain 480/01; MID:21470413; PMID:11586360
 C:Genetics:
 A:Gene: dxt
 C:Superfamily: conserved hypothetical protein HI0807
 C:Keywords: oxidoreductase

Query Match 51.4%; Score 1020; DB 2; Length 398;
 Best Local Similarity 55.0%; Pred. No. 2.7e-69;
 Matches 219; Conservative 51; Mismatches 120; Indels 8; Gaps 3;

QY 1 MKGICILGATGSGTLDVVAHDPKRYQVVALTANGNIDALYECCLAHHPYAVVMS 60
 DB 1 MKGICILGATGSGTLDVVAHDPKRYQVVALTANGNIDALYECCLAHHPYAVVMS 60
 QY 61 KVAEFKQRIASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAARAGKT 120
 DB 61 KVAEFKQRIASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAARAGKT 120
 QY 61 SAQKIMQOHG-SKTEVLSSQACERAAALDEGVHVAIVGAGLPTLAARAGKT 119
 DB 61 SAQKIMQOHG-SKTEVLSSQACERAAALDEGVHVAIVGAGLPTLAARAGKT 119
 QY 121 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPM-----AGYTPGHTAKO 174
 DB 121 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPM-----AGYTPGHTAKO 174
 QY 120 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPM-----AGYTPGHTAKO 178
 DB 120 VLLANKALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCPM-----AGYTPGHTAKO 178
 QY 175 RILITAGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 234
 DB 175 RILITAGSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 234
 QY 179 STIILTSQSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 238
 DB 179 STIILTSQSGPRRTPIETLSVTPDQVAHDKMGRKISVDSATMANKGLELE 238
 QY 235 FNMEDPQIEVVIHPOSIHSMVDYVDGSLAQMGKNDKRTPIAHAMAFPERFDSGAPLD 294
 DB 235 FNMEDPQIEVVIHPOSIHSMVDYVDGSLAQMGKNDKRTPIAHAMAFPERFDSGAPLD 294
 QY 239 FNASARQMEVLIHPOSVIHSVRKYGQSVLAQLGEPDKRTPIAHAMAFPERFDSGAPLD 298
 DB 239 FNASARQMEVLIHPOSVIHSVRKYGQSVLAQLGEPDKRTPIAHAMAFPERFDSGAPLD 298
 QY 295 IFEVGHMDPEKDLRFPCRLRAYEAIKSGGIMPTVLAANAIYEAFLNEBVKFTDLAV 354
 DB 295 IFEVGHMDPEKDLRFPCRLRAYEAIKSGGIMPTVLAANAIYEAFLNEBVKFTDLAV 354

Db 299 FCXGALTTTTPDYQVPCPKAIIDCNAGQAATTALNANETSVAPLDSKIRFTDIEV 358
 355 IIERMAQKPPDAGSLLELVADQDAREVAADIITKL 392
 Db 359 INRTVVEGLLSEPTSEVEVLVDKARADVAQAIVAKL 396

RESULT 10

A64014
 conserved hypothetical protein HI0807 - Haemophilus influenzae (strain Rd KW20)
 C:Species: Haemophilus influenzae
 C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #ext_change 29-Sep-1999
 C:Accession: A64014
 R:Flischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kikness, E.F.; Kellavage, A.; Gockayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.T.; Fuhmann, J.T.; Geophagen, N.S.M. Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MID:95350630; PMID:7542800
 A:Accession: A64014
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-397 <TIGR>
 A:Cross-references: GB:U37263; GB:U42023; MID:91573817; PID:AA022466.1; PID:91573819; M C:Superfamily: conserved hypothetical protein HI0807

Query Match 50.7%; Score 1007; DB 2; Length 397;
 Best Local Similarity 50.6%; Pred. No. 2.6e-68;
 Matches 200; Conservative 76; Mismatches 111; Indels 8; Gaps 3;

Db 2 KGCICIGATGSGTGVSTLDVVAARHPDKQVVALTANGIDALTYGCCLAHREVAIVVWESK 61
 4 QNTVIGSTGSGTGSTLSTVIEENPQTHAFALVGGVNAEMEQCKIRPFAALDDVNA 63
 Db 62 VAEFKORIASPVADIKVLSGSEALQOVATLENVDVMAIVGAAGLPTLAAKAGKT 121
 64 AKILREKILAHNI -PTEVLAGRRALICELAAHPDADQIMASIVGAAGLPTLSAVAKGV 122
 Db 122 LLANKREALVMSGQIFMQAVSDGVALPLPDSSEHNAFQCPMP-----AQYTGHRAKQAR 175
 123 LLANKESLVTQGLFDVAKNYGSKLPVDSSEHNAFQSLPPROKIGFCC-LSBLGVS 181
 Db 176 RLILTAGSGPFRPTPTETLSVTPDQAVAPKMDGRKISVDSATVMNKGLEIIEACILF 235
 182 KIILGSGGPRFTYPIEQFTNITPEQAVAHPMWSGKISVDSATVMNKGLEIIEARMLF 241
 Db 236 NNEPDQIEVVIHPQSIHSMVDYDGSVLAQMGNDPMTPTPIAHAMAPRFDGVAPLDI 295
 242 NNSAEEMEVIIHPOSIIHSMVRVYDGSVITQMGNDPMTPTPIAETVAVPRRTAGVPELDF 301
 Db 296 FEVGHDFEKPDLKRPCLRLAYEALIKSGGIMPTVIANAEIAYEAFINEEYKFTDIAYI 355
 302 FKIKELTYIEPDRNRPNLKALIDAPAQOYATITANNAEINAVQAFIDRQIGFDIAYI 361
 Db 356 IERSMAQKPPDAGSLLELVADQDAREVAADIITKL 390
 362 NSKTERISRPYTIQNDIVLEIDQARERIAKTLR 396

RESULT 11

H82728
 1-deoxy-D-xylulose 5-phosphate reductoisomerase XF1048 [imported] - Xylella fastidiosa
 C:Species: Xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #ext_change 02-Sep-2000
 C:Accession: H82728
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MID:20355717; PMID:10910347
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: H82728

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-396 <STM>
 A:Cross-references: GB:AE003942; GB:AE003849; MID:9105990; PID:AA063858.1; GSPDB:GN001;
 A:Experimental source: strain 945c
 R:Simon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al-
 Bitor, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carver, H.
 as-M, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohne
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzajima, J.P.; Klieger, J.B.; Kuizmae, E.B.; Laigre
 Chad, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, E.
 A:Authors: Martins, B.M.P.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 P.C.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
 Rodrigues, V.; Rosa, A.J.; de M. de Rosa, V.B.; de Sa, R.G.; Santelli, R.V.; Sasaki
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, F.R.; da Silva, A.M.; Silva, W.A.; da Silveira
 M.; Shinako, M.H.; Valada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Ze
 A:Reference number: A59328
 A:Cross-references: annotation
 A:Accession: XF1048
 A:Status: conserved hypothetical protein HI0807

Query Match 47.4%; Score 941; DB 2; Length 396;
 Best Local Similarity 49.2%; Pred. No. 2.4e-63;
 Matches 189; Conservative 71; Mismatches 118; Indels 6; Gaps 2;

Db 1 MKGICIGATGSGTGVSTLDVVAARHPDKQVVALTANGIDALTYGCCLAHREVAIVVWES 60
 5 IRNVAVLATSGTIGAAALDVAAHRRPQVSLAAGQVDDALLCHTTRPDHATGDT 64
 Db 61 KVAEFKORIASPVADIKVLSGSEALQOVATLENVDVMAIVGAAGLPTLAAKAGKT 120
 65 LYTLRDSGNAAGLA -TAYVGEAALAEVASTTCDTVAAIVGAGLHSTLAAABAKR 123
 Db 121 VLLANKREALVMSGQIFMQAVSDGVALPLPDSSEHNAFQCPMPAGYTPGHTAKQARRILL 180
 124 LLANKESLVTAGMLIKREASISGAEIIPDSSEHNAFQCLRSRTD-----VHRIITL 178
 Db 181 ASGGPFRPTPTETLSVTPDQAVAPKMDGRKISVDSATVMNKGLEIIEACILFNMED 240
 179 ASGGPFRPTPTETLSVTPDQAVAPKMDGRKISVDSATVMNKGLEIIEACILFNMED 240
 Db 241 QIEVVIHPQSIHSMVDYDGSVLAQMGNDPMTPTPIAHAMAPRFDGVAPLDIEVGH 300
 229 QIDVAVHPQSIHSMVDYDGSVLAQMGNDPMTPTPIAHAMAPRFDGVAPLDIEVGH 300
 Db 229 QIDVAVHPQSIHSMVDYDGSVLAQMGNDPMTPTPIAHAMAPRFDGVAPLDIEVGH 300
 Db 301 MDPEKDLKRPCLRLAYEALIKSGGIMPTVIANAEIAYEAFINEEYKFTDIAYIERSM 360
 299 LDPEKDLKRPCLRLAYEALIKSGGIMPTVIANAEIAYEAFINEEYKFTDIAYIERSM 360
 Db 361 AQKPPDAGSLLELVADQDAREV 384
 359 TTPRYEADTETITLTVDTETRI 382

RESULT 12

S7633
 hypothetical protein - Synecocystis sp. (strain PCC 6803)
 C:Species: Synecocystis sp.
 A:Accession: PCC 6803
 C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #ext_change 20-Jun-2000
 C:Accession: S76331
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K. Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
 sp.
 A:Reference number: S74322; MID:97061201; PMID:8905231
 A:Accession: S76331
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-394 <KAN>

A:Cross-references: EMBL:D64000; GB:AB001339; NID:g1001484; PIDD:BA10183.1; PIDD:g100155
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Superfamily: conserved hypothetical protein HI0807

Query Match 45.0%; Score 892.5; DB 2; Length 394;
 Best Local Similarity 45.7%; Pred. No. 1,4e-59;
 Matches 183; Conservative 63; Mismatches 126; Indels 13; Gaps 3;

QY 1 MKGICILGATGSGVSTLDVVAHHPDKQVVALTANGNIDALYEQCLAHHPYAVVMS 60
 DB 2 VKRISILGSGTSGIGTDTLIVTHHPAFQVYGLAGNVALLAQVAEFPFVIAIRQAE 61
 QY 61 KVAEFGKRIAASPVADIKTLGSEALQVATLENDVTMAAIVGAGLLPTLAAGA 120
 DB 62 KVEDFAAAVAELTDVQPMVYVGEVGEVYAGDEASVTGIVGAGLLPTLAAGA 121
 QY 121 VLLANKEALVMSGQIFMVAVSDSGAVLPIIDSEHNAIFQCM--PAGYTPGHTAAQARI 177
 DB 122 IALANKETILAGAPVLPVEXKMGVYLLPADSEHSAIFQCLQVPEG-----GLRRI 173
 QY 178 LLTASGAPRRRTPIETLSVTPDQAVAHKMKKISVDSATMANKGHELEACILPM 237
 DB 174 ILTASGAFRRDLPVEELPVTVDALKHPMWSGKITTIDSAITLANKGLEVEIAHYL 233
 QY 236 EPDOIEVVIHPOSIIHSMVDVYDGSVLAQMGNDMPRTPIAHAMAPERFDSGVA 297
 DB 234 DTDHIDIVHPOSIIHSLIEVDSTVLAQMGNDMPRLPLVLSMPERITDMEPLD 293
 QY 298 VGHMDEKDLRFPCLRLAYEAISGSGIMPTVLANAEIYEAFLNEVKFTDIAVIE 357
 DB 294 AGSLSFREPDHDKYPCMQCLAYAGGAGAMPALNNAEQAAVALFOEKISFLDIPRLIE 353
 QY 358 RSMAGKFPDAGS--LEIVLQADQDAR 382
 DB 354 KTCDDLVGGKTPASPDLETLADQMAR 380

RESULT 13

F84957

1-deoxy-D-xylulose 5-phosphate reductoisomerase [imported] - Buchnera sp. (strain APS)

C:Species: Buchnera sp.
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: F84957
 R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000
 A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A

A:Reference number: AB4930; MUID:20445173; PMID:1093077

A:Accession: F84957
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-398 <STO>

A:Cross-references: GB:AP000398; GSPDB:GN00144
 A:Experimental source: strain APS

C:Genetics:

A:Gene: dxr; BU235
 C:Superfamily: conserved hypothetical protein HI0807

Query Match 44.9%; Score 891; DB 2; Length 398;
 Best Local Similarity 45.7%; Pred. No. 1,4e-59;
 Matches 182; Conservative 83; Mismatches 127; Indels 6; Gaps 3;

QY 1 MKGICILGATGSGVSTLDVVAHHPDKQVVALTANGNIDALYEQCLAHHPYAVVMS 60
 DB 1 MKKITTGSGTSGIGTDTLIVTHHPAFQVYGLAGNVALLAQVAEFPFVIAIRQAE 61
 QY 61 KVAEFGKRIAASPVADIKTLGSEALQVATLENDVTMAAIVGAGLLPTLAAGA 120
 DB 61 SAHILKKIKHSKI-NTQVLTGKIDICALAALEETHVISAIVGAGLLPTLSAHAGRT 119
 QY 121 VLLANKEALVMSGQIFMVAVSDSGAVLPIIDSEHNAIFQCMAGYTP--GHTA---KQAR 175
 DB 120 ILLANKEALVMSGQIFMVAVSDSGAVLPIIDSEHNAIFQCMAGYTP--GHTA---KQAR 179

QY 176 RILITAGSGPFRRTPIETLSVTPDQAVAHKMKKISVDSATMANKGHELEACIL 235
 DB 180 HVLTVGGGPPYKSSDLSVTPDQASHPMWLMGKISVDSATMANKGHELEACIL 239
 QY 236 NMEPDQIEVVIHPOSIIHSMVDVYDGSVLAQMGNDMPRTPIAHAMAPERFDSGVA 295
 DB 240 NALASEIKILHPSIIHSMVDVYDGSVLAQMGNDMPRLPLVLSMPERITDMEPLD 299
 QY 296 FEVGHMDEKDLRFPCLRLAYEAISGSGIMPTVLANAEIYEAFLNEVKFTDIAV 355
 DB 300 YKINMLTFEFPDQFQFCLKXIAIDAFSGASMTVLANAEIYEAFLNEVKFTDIAV 359
 QY 356 IERSMAQKFPDAGSLEIVLQADQDAREVARDIKTYV 393
 DB 360 NMEILMSCTSEPKICIDILHIDKRAILANKVSSLI 397

RES T 14

AB2 3

1-deoxy-D-xylulose 5-phosphate reductoisomerase [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AB2349
 R:Kato, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi, S.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

Nature 407, 205-213, 2001
 A:Title: Complete Genome Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. PCC 7120

A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AB2349

A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-399 <STO>
 A:Cross-references: GB:BA000019; PIDD:BA876050.1; PIDD:g1733487; GSPDB:GN00179

A:Experimental source: strain PCC 7120
 A:Genes: alr4351

C:Superfamily: conserved hypothetical protein HI0807

Query Match 43.6%; Score 865; DB 2; Length 399;
 Best Local Similarity 44.8%; Pred. No. 1,3e-57;
 Matches 180; Conservative 76; Mismatches 126; Indels 20; Gaps 5;

QY 1 MKGICILGATGSGVSTLDVVAHHPDKQVVALTANGNIDALYEQCLAHHPYAVVMS 60
 DB 2 VKRISILGSGTSGIGTDTLIVTHHPAFQVYGLAGNVALLAQVAEFPFVIAIRQAE 61
 QY 61 KVAEFGKRIAASPVADIKTLGSEALQVATLENDVTMAAIVGAGLLPTLAAGA 117
 DB 62 KILPAQAALKLDLQPT----LLGEGACVIEVARYGAEITVIGGAGLLPTLAAGA 117
 QY 118 KTYLLANKEALVMSGQIFMVAVSDSGAVLPIIDSEHNAIFQCM--PAGYTPGHTAAQ 174
 DB 118 GKDILANKETILAGAPVLPVEXKMGVYLLPADSEHSAIFQCLQVPEG-----GL 169
 QY 175 RILITAGSGPFRRTPIETLSVTPDQAVAHKMKKISVDSATMANKGHELEACIL 234
 DB 170 KILITAGSGAFRRDMPVDEKLEAVTVSDLKHPMSGKRTVDSATLANKGLEVEIAHYL 229
 QY 235 NMEPDQIEVVIHPOSIIHSMVDVYDGSVLAQMGNDMPRTPIAHAMAPERFDSGVA 294
 DB 230 FGLDVQDIEVVIHPOSIIHSLIEVDSTVLAQMGNDMPRLPLVLSMPERITDMEPLD 289
 QY 295 FEVGHMDEKDLRFPCLRLAYEAISGSGIMPTVLANAEIYEAFLNEVKFTDIAV 354
 DB 290 LVKAGNLFREPDHDKYPCMQCLAYAGGAGAMPALNNAEQAAVALFOEKISFLDIPRLIE 349
 QY 355 IERSMAQKFPDAGS--LEIVLQADQDAREVARDIKTYV 394
 DB 350 CIELVCDHQNDCANPESLDIILAADQASQEVLTATKTLAS 391

RES T 15

H87486 1-deoxy-D-xylulose 5-phosphate reductoisomerase [imported] - *Caulobacter crescentus*

C:\Species: Caulobacter crescentus

C:\Accession: H87486
C:\Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

R; Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.B.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of *Caulobacter crescentus*.
A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: H87486
A-Status: preliminary

A: Molecule type: DNA
A: Position: 1-300 -STO-

A; Cross-references: GB:AE005673; NID:g13423370; PIDN:AAK23892.1; GSPDB:GN00148

C;Genetics:
A;Gene: CCL917

C:Supertamily: conserve

Query Match	43.2%;	Score 857.5;	DB 2
Best Local Similarity	44.6%;	Pred. No. 4.9e-57;	

Matches 175; Conservative 77; Mismatches 125; Indels 15; Gaps 5

QY 4

Db 10 VVLTGSGTIGLSTLSLFEEGAPVQILALTAGRNER

64 EFQRIRASPVADIKVLGSSEALQOVATLENDTVMAIVGAGLIPTLAAAKKQTVLL 123

Dp 70 DLRLAGT--GVEAAGADAVRDPAAM-GADWNSALVGAGLPTVAAARTGAVIAL 129

124 ANKEALVMSGQIFMQAVSDSGAVLLPIDSEHNALPQCMFAGYTPGHTAKQARILLTASG 183

Db 126 ANKSLVCGPALLATAKAGGSVIPDSEHSAIFQLQS-----ECAHRVSRLLITAGS 180

184 GPFRRTPLETLSSVTPDOAVAHPKMDGKRISVDSATMKNKGLELLIACILFNNEPDQIE 243

dbb 181 GPERTWDRKAMARATPRQIAIADPNWMSGAKISYTSATNNKGLRMIFASYSIFATBEDRDVD 240

244 УТИДНОСТИСМТНУМГСУТ НОМСИДНЕРТЯНАМАВЕРЕНСГАДИ ДИРЕВСМОД 303

[illegible]

241 VVIRHQSIVHSINSEI VDS ILAQUSFEDNKAIF IACAR AME FUKLIPWAR KUDLAAIQULIF 300

304 EKPDKKRPPCLAYEAALSGGIMPLVANAANIAYAAFLNEBKVTDIAVLI-----ER-350

Db 301 ESPDVERFPAIGIAREALTLGGGAPAMNMANEVAVAFLDRRIIGELDIAGAVAGILERN 36

Qy 359 -SMAQFKPDDAGSLELVLAQDQAREVARDII 389
| : | : | :
: : : : : :

Db 361 NSLGDLSVAESDAVETAMLIDGSARRIAAEVY 392

Search completed: January 29 2004 15:55:30

```

          201  2007  12.00100
***** completed: Sunday
Job time : 12.5928 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 7.41928 Seconds
(without alignments)
2497.314 Million cell updates/sec

Title: US-09-941-947a-8
Perfect score: 1985
Sequence: 1 MGICLIGATGSGVSTLDV.....LQADQARVARDIITKTLVA 394

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1215.5	61.2	396	1	DXR_PSEAB
2	1091.5	55.0	393	1	DXR_RALSO
3	1062.5	53.0	402	1	DXR_VIBVU
4	1052.5	53.0	394	1	DXR_NEIMA
5	1047.5	52.7	401	1	DXR_VIRPA
6	1045.5	52.6	398	1	DXR_ECOLI
7	1044.5	52.6	398	1	DXR_ECO57
8	1042.5	52.5	405	1	DXR_PASNU
9	1040.5	52.4	394	1	DXR_NEIME
10	1034.5	52.1	398	1	DXR_SALTY
11	1027.5	51.7	398	1	DXR_SALTY
12	1023.5	51.5	402	1	DXR_VIRCH
13	1020.5	51.4	398	1	DXR_YERPE
14	1007.5	50.7	397	1	DXR_HAERI
15	999.5	50.4	398	1	DXR_ZYMOO
16	977.5	49.2	396	1	DXR_XANAC
17	973.5	49.0	396	1	DXR_XANCP
18	941.5	47.4	396	1	DXR_XYLPB
19	892.5	45.0	394	1	DXR_SYNV3
20	892.5	44.9	398	1	DXR_BUCAP
21	891.5	44.9	398	1	DXR_BUCAT
22	887.5	44.7	394	1	DXR_CLOPB
23	886.5	44.6	402	1	DXR_STIME
24	868.5	43.7	382	1	DXR_BACHD
25	865.5	43.6	399	1	DXR_ANNAP
26	857.5	43.2	399	1	DXR_CAUCR
27	854.5	43.0	385	1	DXR_CLOAB
28	854.5	42.8	391	1	DXR_RHIME
29	849.5	42.6	387	1	DXR_AGRBT
30	837.5	42.2	382	1	DXR_THERN
31	818.5	41.2	397	1	DXR_WIGER
32	818.5	41.2	477	1	DXR_ARATH
33	816.5	41.1	380	1	DXR_LISMO

ALIGNMENTS

Result ID	Query	Match	DB ID	Description
1	DXR_PSEAB	396 AA	1	DXR_PSEAB
2	DXR_RALSO	393 AA	1	DXR_RALSO
3	DXR_VIBVU	402 AA	1	DXR_VIBVU
4	DXR_NEIMA	394 AA	1	DXR_NEIMA
5	DXR_VIRPA	401 AA	1	DXR_VIRPA
6	DXR_ECOLI	398 AA	1	DXR_ECOLI
7	DXR_ECO57	398 AA	1	DXR_ECO57
8	DXR_PASNU	405 AA	1	DXR_PASNU
9	DXR_NEIME	394 AA	1	DXR_NEIME
10	DXR_SALTY	398 AA	1	DXR_SALTY
11	DXR_SALTY	398 AA	1	DXR_SALTY
12	DXR_VIRCH	402 AA	1	DXR_VIRCH
13	DXR_YERPE	398 AA	1	DXR_YERPE
14	DXR_HAERI	397 AA	1	DXR_HAERI
15	DXR_ZYMOO	398 AA	1	DXR_ZYMOO
16	DXR_XANAC	396 AA	1	DXR_XANAC
17	DXR_XANCP	396 AA	1	DXR_XANCP
18	DXR_XYLPB	396 AA	1	DXR_XYLPB
19	DXR_SYNV3	394 AA	1	DXR_SYNV3
20	DXR_BUCAP	398 AA	1	DXR_BUCAP
21	DXR_BUCAT	398 AA	1	DXR_BUCAT
22	DXR_CLOPB	402 AA	1	DXR_CLOPB
23	DXR_STIME	402 AA	1	DXR_STIME
24	DXR_BACHD	382 AA	1	DXR_BACHD
25	DXR_ANNAP	399 AA	1	DXR_ANNAP
26	DXR_CAUCR	399 AA	1	DXR_CAUCR
27	DXR_CLOAB	385 AA	1	DXR_CLOAB
28	DXR_RHIME	391 AA	1	DXR_RHIME
29	DXR_AGRBT	387 AA	1	DXR_AGRBT
30	DXR_THERN	382 AA	1	DXR_THERN
31	DXR_WIGER	397 AA	1	DXR_WIGER
32	DXR_ARATH	477 AA	1	DXR_ARATH
33	DXR_LISMO	380 AA	1	DXR_LISMO

DR EMBL: AE004785; AAC07038.1; --
 DR PIR: E83188; E83188.
 DR HAMAP: MF_00183; 1.
 DR InterPro: IPR003821; DXR_reductoisom.
 DR Pfam: PF02670; DXR_reductoisom; 1.
 DR TIGRFAMs: TIGR00243; Dxr; 1.
 DR Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT NP BIND 10 NADPH (POTENTIAL).
 SQ SEQUENCE 396 AA; 42508 MW; 2ECB9D8A94A9CAD7 CRC64;

Query Match 61.2%; Score 1215.5; DB 1; Length 396;
 Best Local Similarity 64.1%; Pred. No. 9,3e-80;
 Matches 245; Conservative 48; Mismatches 88; Indels 1; Gaps 1;

QY 4 IILGATGSGVSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESVA 63
 DB 7 ISVLGATGSGVSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESVA 66
 QY 64 EFKRIASPVADIKYLSGSEALQVATLENDVTMAAIVGAGLLPTLAAKAKGTVL 123
 DB 67 ALQGSIAAAGIR-TEVLFGEQALCEVASAPEYDMYMAIVGAGLLPTLAAVEAGKRVTL 125
 QY 124 AKKALVMSGQIFMQAVSDGAVLLPDSSENAIFQCPAGTPTGHTAKQARRILLTASG 183
 DB 126 AKKALVMSGALFQAVRSQAVLLPDSSENAIFQSPRNADGLERVGVRILLTASG 185
 QY 184 GFFRRPTIETLSSVTPDQAVAHKPMGKRTISVDSATMANKGLLEIACLLFNNEPDQIE 243
 DB 186 GFFRRPTIETLSSVTPDQAVAHKPMGKRTISVDSATMANKGLLEIACLLFNNEPDQIE 245
 QY 244 VVIHQSIHSWVDYDGSVLAQMGKPMKRTIPALHMAWPERFDSGVAFLDIFEGANDF 303
 DB 246 VVIHQSIHSWVDYDGSVLAQMGKPMKRTIPALHMAWPERFDSGVAFLDIFEGANDF 305
 QY 304 EKPDNRPFCLATLAEALKSGGIMPTVLANNEIAVEAFLEVEFTDIAVIERKMAQF 363
 DB 306 QRPDNRPFCLATLAEALKSGGIMPTVLANNEIAVEAFLEVEFTDIAVIERKMAQF 365
 QY 364 KPDAGSLLEVLQADQAREVA 385
 DB 366 AVTAVESLDQVLAADRRARSAVA 387

RESULT 2

DXR_PALSO STANDARD; PRT; 393 AA.

AC Q8XZ15;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).
 GN DXR OR RSC1410 OR RS05282.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 CC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11030;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salenouat M., Genin S., Artiguenave F., Gouzy J., Mangenot S., Arlat M., Billault A., Brothier P., Camus J.C., Catolico L., Chaudier M., Choisme N., Claudel-Renard C., Cunne S., Demange N., Gaspin C., Lavie M., Moisan A., Robert C., Sautin W., Schlex T., Sigler P., Trebault P., Whalen M., Wincker P., Levy M., Weissbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502 (2002).
 CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction of 1-deoxy-D-xyulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol 4-phosphate (MEP) (by similarity).
 CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) = 1-deoxy-D-xyulose 5-phosphate + NADPH.

CC = 1-deoxy-D-xyulose 5-phosphate + NADPH.
 CC -1- PATHWAY: Nornemalonate terpenoid biosynthesis pathway; second step

CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch>).

DR EMBL: AE004785; CAD15112.1; --
 DR HAMAP: MF_00183; 1.
 DR InterPro: IPR003821; DXR_reductoisom.
 DR Pfam: PF02670; DXR_reductoisom; 1.
 DR TIGRFAMs: TIGR00243; Dxr; 1.
 DR Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT NP BIND 7 NADPH (POTENTIAL).
 SQ SEQUENCE 393 AA; 41763 MW; 5A955A7C8F32E9D CRC64;

Query Match 55.0%; Score 1091.5; DB 1; Length 393;
 Best Local Similarity 58.2%; Pred. No. 6.7e-71;
 Matches 227; Conservative 49; Mismatches 109; Indels 5; Gaps 3;

QY 1 MKGICITGATGSGVSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVES 60
 DB 1 MMRITVLGATGSGVSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVES 60
 QY 61 KMAEKKRIASPVADIKYLSGSEALQVATLENDVTMAAIVGAGLLPTLAAKAKGT 120
 DB 61 KMAEKKRIASPVADIKYLSGSEALQVATLENDVTMAAIVGAGLLPTLAAKAKGT 119
 QY 121 VILANKALVMSGQIFMQAVSDGAVLLPDSSENAIFQCPAGTPTGHTAKQARRILLTASG 179
 DB 120 VILANKALVMSGALFQAVRSQAVLLPDSSENAIFQSPRNADGLERVGVRILLTASG 176
 QY 180 TASGFFRRPTIETLSSVTPDQAVAHKPMGKRTISVDSATMANKGLLEIACLLFNNEPDQIE 239
 DB 177 TASGFFRRPTIETLSSVTPDQAVAHKPMGKRTISVDSATMANKGLLEIACLLFNNEPDQIE 236
 QY 240 DQLEVVHPOSIIHSWVDYDGSVLAQMGKPMKRTIPALHMAWPERFDSGVAFLDIFEGANDF 299
 DB 237 ERIEVLTHPOSIIHSWVDYDGSVLAQMGKPMKRTIPALHMAWPERFDSGVAFLDIFEGANDF 296
 QY 300 HMDPEKDLRPFCLATLAEALKSGGIMPTVLANNEIAVEAFLEVEFTDIAVIERKMAQF 359
 DB 297 TLAEADDLRPFCLATLAEALKSGGIMPTVLANNEIAVEAFLEVEFTDIAVIERKMAQF 356
 QY 360 MAQKPDAGSLLEVLQADQAREVA 389
 DB 357 LARTALGPADSLDTPVFAADQARRARHYI 386

RESL T 3

DXR_PALSO STANDARD; PRT; 402 AA.

AC Q8DBF5;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).
 GN DXR OR VY11866
 OS Vibrio vulnificus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=672;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMCP6;
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,

DB 365 FSDGIGDGLLADQARTRARARITGLT 393

RESULT 5
DXR_VIBPA STANDARD: PRT: 401 AA.

AC 087ME3;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).
GN DXR OR VP2312.
OS *Vibrio parahaemolyticus*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K., Iijima Y., Kajima M., Nakano M., Yamashita A., Kubota Y., Kimura S., Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*.";
RL Lancet 361:743-749 (2003).
CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction of 1-deoxy-D-xyulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol 4-phosphate (MEP) (by similarity).
CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) = 1-deoxy-D-xyulose 5-phosphate + NADPH.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second step.
CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: AF005081; BAC60575.1; -.
DR HAMAB: MF 00183; -1.
KW Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
PT NP BIND 14 NADPH (POTENTIAL).
SQ SEQUENCE 401 AA: 43109 MW: 470537.1A14305BD0 CRC64;

Query Match 52.7%; Score 1047; DB 1; Length 401;
Best local similarity 54.3%; Pred. No. 1e-67;
Matches 222; Conservative 57; Mismatches 100; Indels 30; Gaps 5;

QY 1 MKGICIGANGSGVSTLVNVAHPRKXVVALTAAGNIDALYEQGLAHPRVAVVMS 60
DB 1 MQLCTLLGAGSGVSTLVNVAHPRKXVVALTAAGNIDALYEQGLAHPRVAVVMS 60
QY 61 KVAEFKRI-AASPVADIKVLGSSEALQOVATLENDYVVAALVGAAGLLPTLAACK 119
DB 61 AAVALQSEIHTTSP-NTETVLGCVADALCVAALEEDSVAAALVGAAGLLPTLAACK 118
QY 120 TYLANKEALVMSGOIFMAVSDSGAVLLPISSEHNAIQCPAGTTPHTAKQAR---- 175
DB 120 TYLANKEALVMSGOIFMAVSDSGAVLLPISSEHNAIQCPAGTTPHTAKQAR---- 175
QY 119 RYLLANKALVMSGOIFDAVEQYGAELLPVDESEHNAIFQCLP-----QOVVTNIG 169
DB 119 RYLLANKALVMSGOIFDAVEQYGAELLPVDESEHNAIFQCLP-----QOVVTNIG 169
QY 176 -----RIILTASGPFRRPTLETSSVTPDQAVAPFKMDGRKISVDSATMMNG 225
DB 176 -----RIILTASGPFRRPTLETSSVTPDQAVAPFKMDGRKISVDSATMMNG 225
QY 170 RQMLDHEGSIISILLTSGGPFRRVADLADSVTPAALAHPMNMSGPKISVSATMMNG 229
DB 170 RQMLDHEGSIISILLTSGGPFRRVADLADSVTPAALAHPMNMSGPKISVSATMMNG 229
QY 226 LEIIEACLEFNMPEQIEVVIHQSIITHKWDYVDSVLAQNGNPMKRPPIAHANAMPR 285
DB 226 LEIIEACLEFNMPEQIEVVIHQSIITHKWDYVDSVLAQNGNPMKRPPIAHANAMPR 285
QY 230 LEYIEAKWLFNARDQIKVLIHQSVIHSWVQYRDSVLAQNGEPDMATPIALTVSYP 289
DB 230 LEYIEAKWLFNARDQIKVLIHQSVIHSWVQYRDSVLAQNGEPDMATPIALTVSYP 289

QY 286 FDSGVADLDFEYGHMDFEKDILKRPCLALYBAIKSGGIMPTLVNANBIAYEALINE 345
DB 290 VDAVGEKLEDFYQVBEFTFLQPPFARYPCCLALADACBEGOHATTLAANAHEVADAFINN 349
QY 346 EVKFTDIA-----VIERSMAPQKPDGSGELVYLOADQAREVARDIIX 390
DB 350 RLGFTDIARINELVLIKHTASCTEPNANSISLELDNRMERTALAEIR 398

RESI T 6
ID DXR COLI STANDARD: PRT: 398 AA.
AC 24568; P77209;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).
GN DXR OR B0173.
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=K12 / W3110;
RX MEDLINE=98374274; PubMed=9707569;
RA Takahashi S., Kuzuyama T., Matsumoto H., Seto H.;
RT "A 1-deoxy-D-xyulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-D-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:9879-9884 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blatter F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.U., Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX Takemoto K., Mori H., Murayama N., Katsoka K., Yano M., Itoh T., Yamamoto Y., Inokuchi H., Inokuchi H., Hatada B., Fukuda R., Ichihara S., Mizuno T., Makino K., Nakata A., Yura T., Sangei G., Mitsuuchi K.;
RT "Systematic sequencing of the *Escherichia coli* genome: analysis of the 4.0 - 6.0 min (189,987 - 281,416bp) region.";
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S. Duncan M., Allen B., Araujo R., Aparicio A., Chung E., Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurtz O., LaSakari D., Lew H., Lin D., Namath A., Oelner P., Roberts D., Davis R.W.;
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-49 FROM N.A.
RC MEDLINE=93077430; PubMed=1447125;
RX Yamana K., Ogura T., Niki H., Hiraga S.;
RT "Identification and characterization of the *smbA* gene, a suppressor of the *mukB* null mutant of *Escherichia coli*.";
RL J. Bacteriol. 174:7517-7526 (1992).
RN [6]
RP IDENTIFICATION.
RC MEDLINE=96032851; PubMed=7567469;
RX Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C., Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for three gene classes.";

RL Nucleic Acids Res. 23:3554-3562(1995).
RN [7]
RP CHARACTERIZATION
RX MEDLINE=20123893; PubMed=10631325;
RA Radykewicz T., Rohdich F., Mungstntaweekul J., Hertz S., Kis K.,
RT Eisenreich W., Bacher A., Zenk M.H., Arigoni D.;
RT "Biosynthesis of terpenoids: 1-deoxy-D-xylulose-5-phosphate
reductoisomerase from Escherichia coli is a class B dehydrogenase.";
RL FEBS Lett. 465:157-160(2000).
RN [8]
RP CHARACTERIZATION AND MUTAGENESIS.
RX MEDLINE=20347905; PubMed=10787409;
RA Kuriyama T., Takahashi S., Takagi M., Seto H.;
RT "Characterization of 1-deoxy-D-xylulose 5-phosphate reductoisomerase,
an enzyme involved in isopentenyl diphosphate biosynthesis, and
RT identification of its catalytic amino acid residues.";
RL J. Biol. Chem. 275:19928-19932(2000).
CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction
of 1-deoxy-D-xylulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol
4-phosphate (MEP).
CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)
= 1-deoxy-D-xylulose 5-phosphate + NADPH.
CC -1- COFACTOR: REQUIRES A DIVALENT CATION; MAGNESIUM; MANGANESE OR
COBALT.
CC -1- ENZYME REGULATION: INHIBITED BY FOSMIDOMYCIN.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second
step.
CC -1- SUBUNIT: Homotetramer.
CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/sib.ch>).
CC or send an email to license@sib-sib.ch).

DR EMBL: AB013300; BAA32426.1; -
DR EMBL: AE000126; AAC73284.1; -
DR EMBL: D83536; BAA7848.1; -
DR EMBL: U70214; AAB08602.1; -
DR EMBL: D13334; -; NOT ANNOTATED_CDS.
DR PIR: B64741; E64741.
DR PDB: 1K5H; 27-FEB-02.
DR Ecogene: BG12715; dxr.
DR HAMAP: MF_00183; -; 1
DR InterPro: IPR003821; DXR_reductoisomase.
DR Pfam: PF02670; DXR_reductoisom_1.
DR TricorePfam: TIGR00243; Dxr; 1.
KW isoprene biosynthesis; Oxidoreductase; NADP; Magnesium; Manganese;
KW Cobalt; Complete proteome; 3D-structure.
KW NP BIND 7 14 NADPH (POTENTIAL).
FT MUTAGEN 153 14 H->D: DECREASE IN ACTIVITY.
FT MUTAGEN 153 153 H->Q: INCREASE IN KM.
FT MUTAGEN 209 209 H->Q: INCREASE IN KM.
FT MUTAGEN 231 231 E->K: DECREASE IN ACTIVITY.
FT MUTAGEN 257 257 H->Q: INCREASE IN KM.
FT CONFLICT 277 284 RPIPIHTM -> VROLPPPW (IN REF. 3).
SQ SEQUENCE 398 AA; 43388 MW; 9B5326834AFF1207 CRC64;

Query Match 52.6%; Score 1045; DB 1; Length 398;
Best Local Similarity 54.9%; Pred No. 1.4e-67;
Matches 219; Conservative 56; Mismatches 118; Indels 6; Gaps 3;

QY 1 MKGICIIATGSGIVSTLDVVAHEDKTYQVVALTANGNIDALYECCIAHHEPVAVVMS 60
DB 1 MKQLTILSTGSIIGSTLDVVAHNEHFRRVVALVAGKVTWVEECLEFSPRYAMDSEA 60
QY 1 KVAEKKQIIAASPVAIDIKVLSGSESLAQVATLENDTVMAALVGAAGLLPTLAAKAKT 120
DB 61 S-AKLLKTKMLQQGQSGSTEVLSGQQAACDPAALIEDVDQVMAALVGAAGLLPTLARIAGKT 119

```

175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```


RA Eisen U.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gaim M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Cline H., Clark E.B.,
RA Cotton M.D., Ulfersback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Masihami V., Piazza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., McKon B.R., Knapoli R., Venter J.C.,
RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain
MC58." ;
RT MC58." ;
CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction
of 1-deoxy-D-xylose-5-phosphate (DXP) to 2-C-methyl-D-erythritol
4-phosphate (MEP) (By similarity).
CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)
= 1-deoxy-D-xylose 5-phosphate + NADPH.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second
step.
CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AEO02375; AAF4064.1; -.
CC DR TIGR: A81229; A81229.
CC DR TIGR: NM00184; -.
CC DR HAMAP: MF_00183; -; 1.
CC DR InterPro: IPR003821; DXP reductoisomerase.
CC DR Pfam: PF02670; DXP reductoisom; 1.
CC DR TIGRfam: TIGR00243; Dxr; 1.
CC KM Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
CC NP BIND 9 NADPH (POTENTIAL).
CC SQ SEQUENCE 394 AA; 41920 MW; AAB539785B87BD2E CRC64;

CC Query Match 52.4%; Score 1040.5; DB 1; Length 394;
CC Best Local Similarity 55.3%; Pred. No. 3e-67;
CC Matches 215; Conservative 47; Mismatches 126; Indels 1; Gaps 1;

QY 4 ICLIGATSGVSTLVDVAHPDKYQVVALTANGNIDALYEQCLAHHPREYAVVWESKVA 63
DB 6 LTLGSGSIGESTLDVVSHPREKRVFALAGHQAQVLAQOCTFHPREYAVVADAEHAA 65
QY 64 EFKORIASVADIKVTSSEALQCVATLENVDTWMAITGAAGLPTLAAAGKTVLL 123
DB 66 RLRLALRRDGTATQVHLGAQALVDVASADVSGWCAIVGAVGLPSALAAQKGTIYL 124
QY 124 ANKEALVMSGQIFMOAVSDSGAVLPIFDSSEHNAIFQCPAGYTPGHTAKQARRILLTASG 183
DB 125 ANKETTVSGALPMTARANGAAVLPDSEHNAIFQVLPDYAGRLNEHIAIILLTASG 184
QY 184 GPFRRPIETLSVTPDOAVAHHPKMGKISVDSATMANKGLLEIAHMLFCCPDKLE 243
DB 185 GPFLLTADIAFTDRITPAQAVGHPMKRGKISVDSATMANKGLLEIAHMLFCCPDKLE 244
QY 244 VVLIHPOSIIHSMDVYVGSVLAQWGNPDNRPPIAHAAAMPFDSGVAFDIEVGMDF 303
DB 245 VVLIHPOSIIHSMDVYVGSVLAQWGNPDNRPPIAHAAAMPFDSGVAFDIEVGMDF 304
QY 304 EKPLKRPFCILRLAYEALIKSGGIMPTVLANEIAVEAFINEEVKFTDIAVITERSMAQF 363
DB 305 QKPFDFPFCILRLAYEALIKSGGIMPTVLANEIAVEAFINEEVKFTDIAVITERSMAQF 364
QY 364 KPDAGSLVLTQADODAREVADIIKTL 392
DB 365 FSDGSLVLTQADODAREVADIIKTL 393

ID XR_SALTY STANDARD; PRT; 398 AA.
AC 382RP3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 8-FEB-2003 (Rel. 41, Last sequence update)
DT 8-FEB-2003 (Rel. 41, Last annotation update)
DE 1-deoxy-D-xylose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP
reductoisomerase) (1-deoxyxylose-5-phosphate reductoisomerase).
GN XR OR SMO220.
OS Salmonella typhimurium.
OC Acteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX CBI_TaxID=602;
RN 11
RP SEQUENCE FROM N.A.
RC TRAIIN=U2 / SGSC1412 / ATCC 700720;
RC IEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Yeard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
RA Yeaman E., Sun H., Florea L., Miller W., Stonking T., Nhan M.,
RA Asterston R., Wilson R.K.,
RT Complete genome sequence of *Salmonella enterica* serovar Typhimurium
RT 12." ;
RT Nature 413:852-856(2001).
CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction
of 1-deoxy-D-xylose-5-phosphate (DXP) to 2-C-methyl-D-erythritol
4-phosphate (MEP) (By similarity).
CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)
= 1-deoxy-D-xylose 5-phosphate + NADPH.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second
step.
CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AEO08705; AAL19184.1; -.
CC DR Tygene; SG7777; dxr.
CC DR HAMAP: MF_00183; -; 1.
CC DR InterPro: IPR003821; DXP reductoisomerase.
CC DR Pfam: PF02670; DXP reductoisom; 1.
CC DR TIGRfam: TIGR00243; Dxr; 1.
CC KM Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
CC NP BIND 14 NADPH (POTENTIAL).
CC FT BIND 7 14
CC SQ SEQUENCE 398 AA; 43352 MW; CB720D959CDFBFC CRC64;

CC Query Match 52.1%; Score 1034; DB 1; Length 398;
CC Best Local Similarity 55.8%; Pred. No. 8.e-67;
CC Matches 223; Conservative 48; Mismatches 105; Indels 24; Gaps 6;

QY 1 MKGICILATGSGVSTLVDVAHPDKYQVVALTANGNIDALYEQCLAHHPREYAVVWESKVA 60
DB 1 MKGICILATGSGVSTLVDVAHPDKYQVVALTANGNIDALYEQCLAHHPREYAVVWESKVA 60
QY 60 LTLGSGSIGESTLDVVSHPREKRVFALAGHQAQVLAQOCTFHPREYAVVADAEHAA 119
DB 61 LTLGSGSIGESTLDVVSHPREKRVFALAGHQAQVLAQOCTFHPREYAVVADAEHAA 119
QY 121 VVLIHPOSIIHSMDVYVGSVLAQWGNPDNRPPIAHAAAMPFDSGVAFDIEVGMDF 170
DB 122 VVLIHPOSIIHSMDVYVGSVLAQWGNPDNRPPIAHAAAMPFDSGVAFDIEVGMDF 179
QY 171 ANKARIIILTASGPFRRPIETLSVTPDOAVAHHPKMGKISVDSATMANKGLLEIE 230
DB 180 S-----ILTLGSGGPFRRPIETLSVTPDOAVAHHPKMGKISVDSATMANKGLLEIE 234
QY 231 ACLEFMEPDQIBVVIHQSIHSMDVYVGSVLAQWGNPDNRPPIAHAAAMPFDSGVA 290

RESULT 10
DXR_SALTY

DB 235 ARMLFNASARQMEVLIHPOSVTHSMVRYQDSVLAQLGSPDKRPTIAHTAMPKRVTSGA 294
 QY 231 APLDIFEVGHNDPEKPKRPPCLRLAYEAIKSGIMPTVLNANEIAYEAPLNEEVKT 350
 DB 235 QPLDCKLSALTFFSAPDYQRYPCILKAMEAFEGQAATTALNANRITVAFLAQRIPT 354
 QY 351 DIA----VITERSMAQFKEDDAGSLVLIQADQDAREVAR 386
 DB 355 DIAGINLAVALERMDIQ---EPASVEDVLYQVDAIAREVAR 390

RESULT 11

DXR_SALTI STANDARD; PRT; 398 AA.
 AC 0828A6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).
 GN DXR OR STY0243 OR T0221.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
 OX NCBI_TaxId=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RA MEDLINE=21534947; PubMed=11677608; Thomson N.R., Pickard D., Main J., Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Basham D., Brooks K., Chillingworth T., Conger P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gea P., Parry C., Quail M., Rutherford K., Simmonds W., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typh CT18.";
 RT Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RA MEDLINE=22531367; PubMed=12644504; Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
 RA "Comparative genomics of Salmonella enterica serovar Typh strains Ty2 and CT18.";
 RT J. Bacteriol. 185:2330-2337(2003).
 RN [3]
 RP FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction of 1-deoxy-D-xyulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol 4-phosphate (MEP) (By similarity).
 CC 4-CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) = 1-deoxy-D-xyulose 5-phosphate + NADPH.
 CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; second step.
 CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AL627266; CAD08678.1;
 DR HAMM: AEO16834; AA067951.1;
 DR HAMM: MF_00183; -1.1
 DR InterPro: IPR003821; DXP_reductoisomerase.
 DR Pfam: PF02670; DXP_reductoisom; 1.
 DR TIGRPFAM: TIGR00243; Dxr; 1.
 KM Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.

FT NP BIND 7 14 NADPH (POTENTIAL).
 SO SEQUENCE 398 AA; 43328 MW; 89BA627582BC0B5 CRC64;

Q: zy Match 51.7%; Score 1027; DB 1; Length 398;
 Be t Local Similarity 55.8%; Pred. No. 2.8e-66;
 M: ches 223; Conservative 49; Mismatches 104; Indels 24; Gaps 7;

QY 1 MKGICIGANGSGIVSTLDVAVARHPDKRYQVVALTANGNDALYEQCLAHHPYAVVMES 60
 DB 1 MKGLTIGSGTSGICSTLDVYHNPDSFRYIALVAGNVARMNEQCLFSPRYAVMDTS 60
 QY 61 KVAFKQRIASPVADIKVLSSSEALQVATLENDVTMAIYGAAGLPTLAARKGT 120
 DB 61 SABQLKTMLOOHG-SRTEVLSCQAACEAALDEGVHMAIYGAAGLPTLAIRAGKT 119
 QY 121 VILANKALVMSGQIFMQAVSDGAVLIPIDSEHNAIFQGP-----AGTP-----GHT 170
 DB 120 ILLANKESLVTCGLFDEYKRSYARLLPYDSHNAIFQSLPOSIGHNLGYADLEQNGVT 179
 QY 171 AKQARRIILTAAGGPFRRPTIETLSVTPDOAVAHPRMDGKRKISVDSATMNGLEIE 230
 DB 180 S-----ILTSGSGPFRETFMCDLAAMTPQACGHPYMSGKRKISVDSATMNGLEIE 234
 QY 231 ACILFPNNEPDQIEVTHPOSTIHSMDVYNGSVLAQMGNDMPRIAHAMPERFSGV 290
 DB 235 ARMLFNASARQMEVLIHPOSVTHSMVRYQDSVLAQLGSPDKRPTIAHTAMPKRVTSGA 294
 QY 291 APLDIFEVGHNDPEKPKRPPCLRLAYEAIKSGIMPTVLNANEIAYEAPLNEEVKT 350
 DB 295 QPLDCKLSALTFFSAPDYQRYPCILKAMEAFEGQAATTALNANRITVAFLAQRIPT 354
 QY 351 DIA----VITERSMAQFKEDDAGSLVLIQADQDAREVAR 386
 DB 355 DIAGINLAVALERMDIQ---EPASVEDVLYQVDAIAREVAR 390

REST T 12

DXR_IIBCH STANDARD; PRT; 402 AA.

AC 09KEV8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).
 GN DXR OR VC2254.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.
 OX NCBI_TaxId=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RA MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Zmralava M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Frazer C.M.;
 RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.";
 RT Nature 406:447-483(2000).
 RN [2]
 RP FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction of 1-deoxy-D-xyulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol 4-phosphate (MEP) (By similarity).
 CC 4-CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) = 1-deoxy-D-xyulose 5-phosphate + NADPH.
 CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; second step.
 CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: AE004297; AAF95398.1; --
 CC DR PIR: D82099; D82099.
 CC DR TIGR: VC2254; --
 CC DR HAMAP: MF_00183; -- 1.
 CC DR InterPro: IPR003621; DXP_redoxisomerase.
 CC DR Pfam: PF02670; DXP_reductoisom; 1.
 CC DR TIGRFAMs: TIGR00243; Dxr; 1.
 CC DR Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
 CC KW NP_BIND 7 NADPH (POTENTIAL).
 CC FT SEQUENCE 402 AA; 43684 MW; 725DF0BCDC27CE6 CRC64;
 CC SQ

Query Match 51.5%; Score 1023; DB 1; Length 402;
 Best Local Similarity 52.7%; Pred. No. 5.5e-66;
 Matches 214; Conservative 65; Mismatches 103; Indels 24; Gaps 4;

QY 1 MKGICILGATGSGIVSTLDVVARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWES 60
 1 MKKLTILGATGSGIVSTLDVVARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWES 60
 DB 61 KVAEKKRIASPVADIKYLSGSEALQOYATLENDVTMAAIVGAAGLPTLAAKAKKT 120
 61 KVAEKKRIASPVADIKYLSGSEALQOYATLENDVTMAAIVGAAGLPTLAAKAKKT 119
 QY 121 VILANKREALVMSGGIFMQAVSDSGAVLLPIDSEHNALFQCMGATPGHTAKQAR----- 175
 121 VILANKREALVMSGGIFMQAVSDSGAVLLPIDSEHNALFQCMGATPGHTAKQAR----- 175
 DB 120 VILANKREALVMSGGIFMQAVSDSGAVLLPIDSEHNALFQCMGATPGHTAKQAR----- 172
 120 VILANKREALVMSGGIFMQAVSDSGAVLLPIDSEHNALFQCMGATPGHTAKQAR----- 172
 QY 176 -----RILITAGSGGPFRRPTIFLTSSVTPDQAVAHPRKMGKISVDSATMNGKIEL 228
 176 -----RILITAGSGGPFRRPTIFLTSSVTPDQAVAHPRKMGKISVDSATMNGKIEL 228
 DB 173 LSGHGDHILTLTGSGGPFRRPTIFLTSSVTPDQAVAHPRKMGKISVDSATMNGKIEL 232
 173 LSGHGDHILTLTGSGGPFRRPTIFLTSSVTPDQAVAHPRKMGKISVDSATMNGKIEL 232
 QY 229 IEACLENNERPQIEVYHPOSIIHSMVDYVGVLAQMGNDMPTPLAHMAMPERDS 288
 229 IEACLENNERPQIEVYHPOSIIHSMVDYVGVLAQMGNDMPTPLAHMAMPERDS 288
 DB 233 IEAKMLFNTSRQKLKVLHPQSVHSMVOYQDSVIAQIGEDDMTPSTINAIERYVTA 292
 233 IEAKMLFNTSRQKLKVLHPQSVHSMVOYQDSVIAQIGEDDMTPSTINAIERYVTA 292
 QY 289 GVAPIIDPEVGHMDEKDPDLKRPCLRLAYBAIKSGIMPTVLANANEIAYEALFMBEYK 348
 289 GVAPIIDPEVGHMDEKDPDLKRPCLRLAYBAIKSGIMPTVLANANEIAYEALFMBEYK 348
 DB 293 GVPALDEFTLQQLTMEVDFAFYPTCLQALMDACFGVATTSILNANAVADAFLKRIK 352
 293 GVPALDEFTLQQLTMEVDFAFYPTCLQALMDACFGVATTSILNANAVADAFLKRIK 352
 QY 349 FFDIAVITIRSMAGKRPD-----AGSLVYLQADQDAEVARDIK 390
 349 FFDIAVITIRSMAGKRPD-----AGSLVYLQADQDAEVARDIK 390
 DB 353 FFDIALINDQVLKVCATNTQLHCRDLSELTLELDTMARHFAHQVLK 398
 353 FFDIALINDQVLKVCATNTQLHCRDLSELTLELDTMARHFAHQVLK 398

RESULT 13
 DXR_YERPE STANDARD; PRT; 398 AA.
 AC Q82H62;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 1-deoxy-D-xylose 5-phosphate reductoisomerase [EC 1.1.1.267] (DXP
 reductoisomerase) (1-deoxyxylyulose-5-phosphate reductoisomerase).
 GN DXR OR YPO1048 OR Y1311.
 OS Versinia pests.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Versinia.
 OX NCBI_TaxID=632;
 RN (1)
 RC SEQUENCE FROM N.A.
 RP STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parthill U., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,

RA Millingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Atwell T., Hamlin N., Holroyd S., Jagers K., Kariyeh A.V.,
 RA Sather M., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Iamonds M., Skelton J., Stevens K., Whitehead S., Barrill B.G.,
 RA The genome sequence of Versinia pests, the causative agent of plague.";
 RA Nature 413:523-527(2001).
 RN 2)
 RP SEQUENCE FROM N.A.
 RC TRAIN=KIMS / Biovar Mediaevalis;
 RX EDLIN=22137863; PubMed=12142430;
 RA Ang W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Zima N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Shestern J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Cravay S.C., McDonough K.A., Niles M.L., Matson J.S., Blattner F.R.,
 RA Berry R.D.,
 RA The genome sequence of Versinia pests KIM.";
 RL Bacteriol. 184:4601-4611(2002).
 CC FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction
 CC of 1-deoxy-D-xylose-5-phosphate (DXP) to 2-C-methyl-D-erythritol
 CC 4-phosphate (MEP) (By similarity).
 CC CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)
 CC = 1-deoxy-D-xylose 5-phosphate + NADPH.
 CC PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second
 CC step.
 CC SIMILARITY: BELONGS TO THE DXR FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: AJ414146; CAC89890.1; --
 CC DR EBL: AE013914; AAM86681.1; --
 CC DR IR: AG0128; AG0128.
 CC DR HAMAP: MF_00183; -- 1.
 CC DR InterPro: IPR003621; DXP_redoxisomerase.
 CC DR Pfam: PF02670; DXP_reductoisom; 1.
 CC DR TIGRFAMs: TIGR00243; Dxr; 1.
 CC DR Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
 CC KW NP_BIND 7 NADPH (POTENTIAL).
 CC FT SEQUENCE 398 AA; 43115 MW; C9B1FC9E0165D057 CRC64;
 CC SQ

Query Match 51.4%; Score 1020; DB 1; Length 398;
 Best Local Similarity 55.0%; Pred. No. 8.8e-66;
 Matches 219; Conservative 51; Mismatches 120; Indels 8; Gaps 3;

QY 1 MKGICILGATGSGIVSTLDVVARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWES 60
 1 MKKLTILGATGSGIVSTLDVVARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWES 60
 DB 61 KVAEKKRIASPVADIKYLSGSEALQOYATLENDVTMAAIVGAAGLPTLAAKAKKT 120
 61 KVAEKKRIASPVADIKYLSGSEALQOYATLENDVTMAAIVGAAGLPTLAAKAKKT 119
 QY 121 VILANKREALVMSGGIFMQAVSDSGAVLLPIDSEHNALFQCMGATPGHTAKQAR----- 174
 121 VILANKREALVMSGGIFMQAVSDSGAVLLPIDSEHNALFQCMGATPGHTAKQAR----- 174
 DB 120 VILANKREALVMSGGIFMQAVSDSGAVLLPIDSEHNALFQCMGATPGHTAKQAR----- 178
 120 VILANKREALVMSGGIFMQAVSDSGAVLLPIDSEHNALFQCMGATPGHTAKQAR----- 178
 QY 175 RILITAGSGGPFRRPTIFLTSSVTPDQAVAHPRKMGKISVDSATMNGKIELIEACTL 234
 175 RILITAGSGGPFRRPTIFLTSSVTPDQAVAHPRKMGKISVDSATMNGKIELIEACTL 234
 DB 179 SRILITAGSGGPFRRPTIFLTSSVTPDQAVAHPRKMGKISVDSATMNGKIELIEACTL 238
 179 SRILITAGSGGPFRRPTIFLTSSVTPDQAVAHPRKMGKISVDSATMNGKIELIEACTL 238
 QY 235 FNEPDDIIVAVIHQSIIHSMVDYVGVLAQMGNDMPTPLAHMAMPERDSGVAPLD 294
 235 FNEPDDIIVAVIHQSIIHSMVDYVGVLAQMGNDMPTPLAHMAMPERDSGVAPLD 294
 DB 239 FNAAGQIEVYVLFHQSYHSHVRTHDSIIAQMGTPMRTPIAHAMAYPMKVSAGVAPLD 298
 239 FNAAGQIEVYVLFHQSYHSHVRTHDSIIAQMGTPMRTPIAHAMAYPMKVSAGVAPLD 298
 QY 295 IFVYGHDFEKPDLKRPCLRLAYBAIKSGIMPTVLANANEIAYEALFMBEYKTDIIV 354
 295 IFVYGHDFEKPDLKRPCLRLAYBAIKSGIMPTVLANANEIAYEALFMBEYKTDIIV 354
 DB 299 FCKRGALFTTPDYQRYPCIKLAIDACNAGQAATTAANNEISVMAFLDSKIRFDIEV 358
 299 FCKRGALFTTPDYQRYPCIKLAIDACNAGQAATTAANNEISVMAFLDSKIRFDIEV 358

QY 355 IERSMAQFKPDGSLVLYQADQDAREVARDIITL 392
 DB 359 INRTVEGLLSEPTSEVEVLVDKRDVAAQVIATKL 396

RESULT 14
 DNR_HAEMIN STANDARD; PRT: 397 AA.
 AC P44055;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP
 reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).
 GN DNR OR H10807;
 OS Haemophilus influenzae;
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 CC NCBI_TaxID=727;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KM20 / ATCC 51907;
 RX MEDLINE=9530630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Keriavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shiley R., Liu L.-I., Glodok A., Kelley J.M.,
 RA Weiman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utechtack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
 RA Venter J.C., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd".
 RT Science 263:496-512(1995).
 RL -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction
 of 1-deoxy-D-xyulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol
 4-phosphate (MEP) (By similarity).
 CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)
 = 1-deoxy-D-xyulose 5-phosphate + NADPH.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second
 step.
 CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: U32763; AAC22466.1; -
 DR PIR: A64014; A64014.
 DR TIGR: H10807; -
 DR HAMAP: MF_00183; -; 1.
 DR InterPro: IPR003821; DXP_reductoisomerase.
 DR Pfam: PF02670; DXP_reductoisom; 1.
 DR TIGRFAMs: TIGR00243; Dxr; 1.
 KM Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT NP BIND 9 NADPH (POTENTIAL).
 SQ SEQUENCE 397 AA; 43667 MW; A659B24CCGALCEAA CR654;

Query Match 50.7%; Score 1007; DB 1; Length 397;
 Best Local Similarity 50.6%; Pseq. No. 7,5e-65;
 Matches 200; Conservative 76; Mismatches 111; Indels 8; Gaps 3;

QY 2 KSCICIGATGSGIVGTLVAVARPPYQVVALTANGINDATVBOGLAHPEYAVVWMSK 61
 DB 4 QMIVLIGSGISGKSTLSTIENNPKTHAFALVGRVNEAMRQICRPHRALDDVNA 63
 QY 62 VAEFKRIASPVADIKVLSSGSEALQGVATLENVDVTWMAIVGAAGLLPTLAARAGTV 121

DB 64 AKIIREULIHHI -PTEVLAGRRALICELAAHPDADQIMASIVGAAGLLPTLSAVKAGRV 122
 QY 122 LLANKELVNSGQIFMOAVSDGVALIPDISSENNATFOCP-----AGTPGHIAQAR 175
 DB 123 LLANKESLVTCGLFTDAVNRYSKLLPVDSHNNATFOSLPPRAQKIGFCP-LSEIGVS 181
 QY 176 RILITAGGPFRRPTPIETLSVTPDQAVAHPKMDGRKISVDSATMMNKKLEIEACILP 235
 DB 182 KIITGSGGPFRTPTLPQFINITPEQAVAHPMNSMCKKISVDSATMMNKKLEIEANMLP 241
 QY 236 NMEPDQIEVVIHQSIHNSVVDYDGSVLAQKAPDKRTPIAHANAPPEFDGVALPDI 295
 DB 242 NASAEKREVIHQSIHNSVRYDGSVITQKAPDKRTPIAETMAVPHRTFAGVEPLDF 301
 QY 296 FEVGHMDPEKDLKRPCLAYEATKSGGIMPTVLAANEIVAEVAFINEVKPTDIATV 355
 DB 302 FKIELFTIPDFNRPNLKLATIDAPAGAYATTAMPAANEIVAEVAFINQIGFMADI 361
 QY 356 IERSMAQFKPDGSLVLYQADQDAREVARDIITL 390
 DB 362 NSKTERISPTTIQINDVLEIDAQAREIAKTILR 396

REST T 15
 DNR_XMO STANDARD; PRT: 388 AA.
 AC 09X5P2; O9RI9;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP
 reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).
 GN DXR.
 OS Zymomonas mobilis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
 OC Sphingomonadaceae; Zymomonas.
 CC NCBI_TaxID=542;
 RN (1)
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RC STRAIN=ATCC 31821 / ZM4 / CP4;
 RX MEDLINE=20461176; PubMed=11004410;
 RA Grollie S., Bringer-Meyer S., Sahn H.;
 RA "Isolation of the dxr gene of Zymomonas mobilis and characterization
 of the 1-deoxy-D-xyulose 5-phosphate reductoisomerase.";
 RT FEBS Microbiol. Lett. 191:131-137(2000).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 31821 / ZM4 / CP4;
 RA Lee H.J., Kang H.S.;
 RA Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 RL -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction
 of 1-deoxy-D-xyulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol
 4-phosphate (MEP).
 CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)
 = 1-deoxy-D-xyulose 5-phosphate + NADPH.
 CC -1- COFACTOR: REQUIRES A DIVALENT CATION; MAGNESIUM, MANGANESE OR
 COBALT.
 CC -1- ENZYME REGULATION: INHIBITED BY ROSMIDOMYCIN.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second
 step.
 CC -1- SUBUNIT: Homodimer.
 CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: AJ250714; CAB60758.1; -
 DR EMBL: AF124757; AA029659.1; -;

```

DR HAMAP; MF 00183; -; 1.
DR InterPro; IPR003821; DXP_reductoisomerase.
DR Pfam; PF02670; DXP_reductoisom; 1.
DR TIGRFAMs; TIGR00243; Dxr; 1.
KW Isoprene biosynthesis; Oxidoreductase; NADP; Magnesium; Manganese;
KW Cobalt.
FT NP BIND 10 17 NADPH (POTENTIAL).
FT CONFLICT 226 226 F -> Y (IN REF. 2).
SQ SEQUENCE 388 AA; 41842 MW; 2E161B712089613F CRC64;

Query Match 50.4%; Score 999.5; DB 1; Length 388;
Best Local Similarity 50.3%; Pred. No. 2.5e-64;
Matches 199; Conservative 74; Mismatches 109; Indels 9; Gaps 3;

QY 4 ICIILATGSGIGVSTLDVVRHHPDKYOVALLTANGNIDALYEQCLAHHPDYAVVMESKVA 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7 VTVLGATGSGIGSTLDLERRNLDRCQVALLTANRVKDLADNAKRTNARAVLATDPSLYN 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 64 EFKORIAASPVADIKVLSGSBAIQVATLENDVTMAAIVGAAGLLPTLAAAKAGTVIL 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 67 DLKEALAGS---SVEAAGADALVEAAWV-GADWTMAALIGCAGLKATLAIKKGKVAL 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 124 ANKEALVNSGQIFMQAVSDSGAVLLPDSNNALFQCMFAGTTPGHITAKQARRILLTASG 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 123 ANKESLVNSAGGLMDAVREHGTLLPVDEHNALFQCF----PHNRDYYVRIIITASG 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 184 GPFRRTPETLSSVTPDQAVAHPKWDMGRKISVDSATMNNKGLLEIACLFNMEPDQIE 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 GPFRTTSLAEMAVTTPERAVQHPNWSMGAKISIDATMNNKGLLEIAFHLFOILEKFE 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 244 VVIFQSIHNSWVDYVDSVLAQNGNPDWRTPIAHMAWPERFDSGVAFLDIEVGHMDF 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 238 ILVHFQSVHNSVVEYLDGSIILAQIGSPDWRTPIGHITLAFKMETPASESLDFTKLRQMDP 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 304 EKPDLKRPFCFLAYEATIKSGIMPTVINAANEIAYEAFLNEVEKFTDIAVIERMAQF 363
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 298 EAPDYERFPALTLMASISIKSGARPAVMAANEIYAAAFIDKKIGPLDIKIVEXTLDHY 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 364 KPDDAGSLELVLAQADQAREVARDIIKTLVA 394
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 358 TPATPSLELDVFAIDNEARIQAAALMESLPA 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: January 29, 2004, 15:50:39
 Job time : 8.41938 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:14 ; Search time 29.3684 Seconds
(without alignments) 3461.579 Million cell updates/sec

Title: US-09-941-947a-8
Perfect score: 1985
Sequence: 1 MKGCIIGATSGISGVSTLDV.....LQADQDAEVARADIIKTLVA 394

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23: *
1: sp. archaea: *
2: sp. bacteria: *
3: sp. fungi: *
4: sp. human: *
5: sp. invertebrate: *
6: sp. mammal: *
7: sp. mhc: *
8: sp. organelle: *
9: sp. phase: *
10: sp. plant: *
11: sp. rodent: *
12: sp. virus: *
13: sp. vertebrate: *
14: sp. unclassified: *
15: sp. viirus: *
16: sp. bacteriap: *
17: sp. archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1062	53.5	402	16	Q8DBF5 Q8DBF5
2	1041.5	52.5	396	16	Q8EGG9 Q8EGG9
3	983	49.5	382	16	Q8KGA3 Q8KGA3
4	870	43.8	413	16	Q8DK30 Q8DK30
5	838.5	42.2	472	10	Q8P64 Q8P64
6	827.5	41.7	473	10	Q8L6C5 Q8L6C5
7	818.5	41.2	397	16	Q8D2G6 Q8D2G6
8	816.5	41.1	475	10	Q947C3 Q947C3
9	810.5	40.8	473	10	Q8W250 Q8W250
10	807.5	40.7	473	10	Q9FTNO Q9FTNO
11	803.5	40.5	474	10	Q9M4F4 Q9M4F4
12	802.5	40.4	472	10	Q9PXZ7 Q9PXZ7
13	735.5	37.1	391	16	Q8T146 Q8T146
14	729.5	36.8	488	5	Q96693 Q96693
15	729.5	36.8	488	5	Q8IKG4 Q8IKG4
16	718	36.2	386	2	Q9AJD7 Q9AJD7

1	672.5	33.9	396	16	Q8G7Y7 Q8G7Y7
2	671	33.8	504	16	Q8PP80 Q8PP80
3	531	26.8	377	16	Q8EW06 Q8EW06
4	495	24.9	204	2	Q8KMT5 Q8KMT5
5	297	15.0	170	10	Q8LNG5 Q8LNG5
6	260.5	13.1	115	2	Q9EX11 Q9EX11
7	152	7.7	94	2	Q87046 Q87046
8	127	6.4	781	16	Q8DP12 Q8DP12
9	115.5	5.8	1564	16	Q8PR45 Q8PR45
10	114	5.7	522	16	Q9KM31 Q9KM31
11	111	5.6	3104	2	Q04846 Q04846
12	110.5	5.6	7525	2	Q9KIR0 Q9KIR0
13	110.5	5.6	367	16	Q92A84 Q92A84
14	109.5	5.5	6146	2	Q9JH5 Q9JH5
15	109.5	5.4	456	16	Q8UC43 Q8UC43
16	106	5.3	486	16	Q8NP76 Q8NP76
17	105.5	5.3	367	16	Q8Y5X9 Q8Y5X9
18	105.5	5.3	475	16	Q9AB89 Q9AB89
19	105.5	5.3	2546	2	Q9A130 Q9A130
20	105.5	5.3	3816	2	Q9KIV3 Q9KIV3
21	105	5.3	276	16	Q8NN06 Q8NN06
22	105	5.3	1096	16	Q8PRK7 Q8PRK7
23	105	5.3	2352	16	Q9HTR8 Q9HTR8
24	104.5	5.3	1163	16	Q92U07 Q92U07
25	104.5	5.3	1163	2	Q56780 Q56780
26	104.5	5.3	9507	2	Q9EM41 Q9EM41
27	104	5.2	405	16	Q98AS0 Q98AS0
28	104	5.2	899	16	Q92273 Q92273
29	103.5	5.2	316	16	Q8G115 Q8G115

ALIGNMENTS

Result No.	Score	Query Match	Length	DB ID	Description
1	1062	53.5	402	16	Q8DBF5 Q8DBF5
2	1041.5	52.5	396	16	Q8EGG9 Q8EGG9
3	983	49.5	382	16	Q8KGA3 Q8KGA3
4	870	43.8	413	16	Q8DK30 Q8DK30
5	838.5	42.2	472	10	Q8P64 Q8P64
6	827.5	41.7	473	10	Q8L6C5 Q8L6C5
7	818.5	41.2	397	16	Q8D2G6 Q8D2G6
8	816.5	41.1	475	10	Q947C3 Q947C3
9	810.5	40.8	473	10	Q8W250 Q8W250
10	807.5	40.7	473	10	Q9FTNO Q9FTNO
11	803.5	40.5	474	10	Q9M4F4 Q9M4F4
12	802.5	40.4	472	10	Q9PXZ7 Q9PXZ7
13	735.5	37.1	391	16	Q8T146 Q8T146
14	729.5	36.8	488	5	Q96693 Q96693
15	729.5	36.8	488	5	Q8IKG4 Q8IKG4
16	718	36.2	386	2	Q9AJD7 Q9AJD7

QY 174 ARIILLTASGGFRRPTPIETLSSVTPDQAVAHPKMDGKISVDSATMANKLEIIEACL 233
 DB 178 VSHILLTGSGFRRPTPIETLSSVTPDQAVAHPKMDGKISVDSATMANKLEIIEACL 237
 QY 234 LFNEMPDQEVITHQSIHSHWVDVYDGLAQKNGPKRPTIAPAMAPERFDSGVAFL 293
 DB 238 LFNARSDQKVIITHQSIHSHWVDVYDGLAQKNGPKRPTIAPAMAPERFDSGVAFL 297
 QY 294 DIFEVGHMDPEKRDLPKPCILAYEAIKSGIMPTVANAIEAVEAFINEVETDIA 353
 DB 298 DFTQVGLTFLQDFERRYPCLALIEACYLQGHATTITLANAEVAAFLAQIKFTDIA 357
 QY 354 -----VIERSMAQFKDDAGSLVYQADQDAREVARDIK 390
 DB 358 RNDVSVLNVQCKQSLASGLDSLELTDPMRTTADDEVYR 398

RESULT 2

08EGG9 PRELIMINARY; PRT; 396 AA.
 ID 08EGG9
 AC 08EGG9; 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.
 GN DXR OR S01635.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadales; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1.
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson M.C.,
 Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 Meyer T., Tsapin A., Scott J., Beaman M., Binkac L., Daugherty S.,
 DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 Vamathevan J., Weidman J., Impraim M., Lee C., Berry K., Lee C.,
 Mueller J., Khouri H., Gill J., Uutterback T.R., McDonald L.A.,
 Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,
 RT Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis.
 RL Nat. Biotechnol. 20:1118-1123 (2002).
 DR EMBL; AB015609; AAN54690.1; -
 DR TIGR; S01635; -
 KW Isomerase; Complete proteome.
 SQ SEQUENCE 396 AA; 42159 MW; EDB1785DDE37A3B7 CRC64;

Query Match 52.5%; Score 1041.5; DB 16; Length 396;
 Best local similarity 55.1%; Pred. No. 2.3e-69;
 Matches 217; Conservative 49; Mismatches 121; Indels 7; Gaps 2;

QY 1 MKGICILGATGSGVSTLVDVVAHPPKQYVALTANGNIDALYEOCLAHPEYAVYMES 60
 DB 1 KGMVTLGATGSGVSTLVDVVAHPPKQYVALTANGNIDALYEOCLAHPEYAVYMES 60
 QY 61 KVAEFKORIASPVADIKVLGSEALQOAVATLENVDTMAALVGAAGLPTLAAGAGT 120
 DB 61 KVAEFKORIASPVADIKVLGSEALQOAVATLENVDTMAALVGAAGLPTLAAGAGT 120
 QY 121 VLLANKKALVMSGGQIFMOAVSDSGAVLLPIDEENALFOCPAGTTPG-----TRAKQAR 175
 DB 121 VLLANKKALVMSGGQIFMOAVSDSGAVLLPIDEENALFOCPAGTTPG-----TRAKQAR 175
 QY 176 RILTFASGPPRRPTPIETLSSVTPDQAVAHPKMDGKISVDSATMANKLEIIEACL 235
 DB 176 RILTFASGPPRRPTPIETLSSVTPDQAVAHPKMDGKISVDSATMANKLEIIEACL 235
 QY 238 LFNARSDQKVIITHQSIHSHWVDVYDGLAQKNGPKRPTIAPAMAPERFDSGVAFL 297
 DB 238 LFNARSDQKVIITHQSIHSHWVDVYDGLAQKNGPKRPTIAPAMAPERFDSGVAFL 297
 QY 294 DIFEVGHMDPEKRDLPKPCILAYEAIKSGIMPTVANAIEAVEAFINEVETDIA 353
 DB 294 DIFEVGHMDPEKRDLPKPCILAYEAIKSGIMPTVANAIEAVEAFINEVETDIA 353
 QY 354 -----VIERSMAQFKDDAGSLVYQADQDAREVARDIK 390
 DB 354 -----VIERSMAQFKDDAGSLVYQADQDAREVARDIK 390

DB 239 MTQDQKLVITHQSIHSHWVDVYDGLAQKNGPKRPTIAPAMAPERFDSGVAFL 298
 QY 296 FEVGHMDPEKRDLPKPCILAYEAIKSGIMPTVANAIEAVEAFINEVETDIA 353
 DB 299 FKVQLSFCEPDFRRPCLALIEACQSQGQBATVLANAEIAVEAFLOCKIGFTHICKI 358
 QY 356 IERSMAQFKDDAGSLVYQADQDAREVARDIK 389
 DB 359 NEDCLISVPKQAMASIEDITLADQRTIARELL 392

REST 1 3

08K43 PRELIMINARY; PRT; 382 AA.
 ID 08K43
 AC 08K43; 01-OCT-2002 (TReMBLrel. 22, Created)
 DT 01-OCT-2002 (TReMBLrel. 22, last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.
 GN DXR OR C70125.
 OS Chlorobium tepidum.
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
 OC Chlorobium.
 OX NCBI_TaxID=1097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TL5 / ATCC 49652 / DSM 12025;
 RX MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
 Dodson R.J., DeBoy R., Gwinn M.L., Nelson M.C., Haft D.H.,
 Hickey B.K., Peterson J.D., Durkin A.S., Kolonay J.F., Yang F.,
 Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
 Vamathevan J., Khouri H., Hansen C.L., Craven M.B., Radune D.,
 Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.,
 RT The complete genome sequence of Chlorobium tepidum TL5, a
 RT photosynthetic, anaerobic, green-sulfur bacterium.
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).
 DR EMBL; AB012791; AAM71373.1; -
 DR TIGR; C70125; -
 DR InterPro; IPR003821; DXR reductoisomerase.
 DR Pfam; PF02670; DXR reductoisom; 1.
 DR TIGRFAMs; TIGR00243; DXR; 1.
 KW Isomerase; Complete proteome.
 SQ SEQUENCE 382 AA; 41343 MW; 4B428B6F36368CA CRC64;

Query Match 49.5%; Score 983; DB 16; Length 382;
 Best local similarity 50.6%; Pred. No. 5e-65;
 Mismatches 106; Indels 12; Gaps 3;

QY 1 MKGICILGATGSGVSTLVDVVAHPPKQYVALTANGNIDALYEOCLAHPEYAVYMES 60
 DB 1 MKGICILGATGSGVSTLVDVVAHPPKQYVALTANGNIDALYEOCLAHPEYAVYMES 60
 QY 61 KVAEFKORIASPVADIKVLGSEALQOAVATLENVDTMAALVGAAGLPTLAAGAGT 118
 DB 61 KVAEFKORIASPVADIKVLGSEALQOAVATLENVDTMAALVGAAGLPTLAAGAGT 118
 QY 119 KTVILANKKALVMSGGQIFMOAVSDSGAVLLPIDEENALFOCPAGTTPGHTAQARIL 178
 DB 119 KTVILANKKALVMSGGQIFMOAVSDSGAVLLPIDEENALFOCPAGTTPGHTAQARIL 178
 QY 179 LTFASGPPRRPTPIETLSSVTPDQAVAHPKMDGKISVDSATMANKLEIIEACL 238
 DB 179 LTFASGPPRRPTPIETLSSVTPDQAVAHPKMDGKISVDSATMANKLEIIEACL 238
 QY 239 GMDPEKRDLPKPCILAYEAIKSGIMPTVANAIEAVEAFINEVETDIA 358
 DB 239 GMDPEKRDLPKPCILAYEAIKSGIMPTVANAIEAVEAFINEVETDIA 358
 QY 359 NEDCLISVPKQAMASIEDITLADQRTIARELL 392
 DB 359 NEDCLISVPKQAMASIEDITLADQRTIARELL 392

QY 359 SMAQFKPDDAGSLBLVLAQADODAREVARDI 389
 DB 351 TMOAHEAWPTITLBRVLAQADKWARAROLI 381

RESULT 4
 Q8DK30 PRELIMINARY; PRT; 413 AA.

AC Q8DK30;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.
 GN TLR1040.
 OS Synchococcus elongatus (Thermosynchococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
 OX NCBI_TaxId=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RA MEDLINE=2225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Katayama A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynchococcus elongatus BP-1.";
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AP005372; BAC08593.1;
 KM Isomerase; Complete proteome.
 SQ SEQUENCE 413 AA; 44148 MW; 72760B1507960644 CRC64;

Query Match 43.8%; Score 870; DB 16; Length 413;
 Best Local Similarity 45.4%; Pred. No. 1.5e-56;
 Matches 184; Conservative 76; Mismatches 123; Indels 22; Gaps 6;

QY 1 MGICILGATGSGVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWMSK 60
 DB 19 VQALNLTSGTSGTGYLDIVAGYDRPRVGLAAGNLERIPQIRQFQPIVSIADPE 78
 QY 61 KVAEFQRIASPVDADIKVLSGSEALQOVATLENVDYMAAIVGAAGLLPTLAARAGKT 120
 DB 79 QLPETLALADLPORP-QVAGEAGIAAAYAGDBVVTGIVGAGVPTIAIKAGD 137
 QY 121 VLLANKKALVMSGQIFMQAVSDSGAVLPIIDSEHNAIQCM---PAGYTPGHTAQARI 177
 DB 138 IALANKETLIAGGPVTLPIQETGYKLEPADESHSAIFQCLQGVPEG-----GIKKI 189
 QY 178 LITASGGPFRRTPIETLSSTVDPQAVAHPKMDGKRIYSDATMNNKGLIELIACLLFNM 237
 DB 190 ILTASGAFRDPVPELQAVTADALKHPNMSGKPIITVDSATLNKGLIEVIAHYLFQM 249
 QY 238 EPDQIEVVIHPQSIHSMVDYDGSVLAQMGNDPMTPIAHAMAMPEDSDVAALDIFE 297
 DB 250 DYDNMEIVIHPOSIIHSLIELODTSTVLAQLGMPMLPLVYLSMPERTPMSSPLDLVK 309
 QY 298 VGHMDPEKEDLRFPCRLAIVBAIKSGIMPTVLANAEIAVEATLNEVEKFTDIAVIER 357
 DB 310 AADLTFRSDHQKPYMGGLAYAGAGAMPVLANAGQAVLPIAEIAIATLEIPRIIE 369
 QY 358 ----RSMQFKPDDAGSLBLVLAQADODAR---EVARDIIXLIVA 394
 DB 370 MWCDRYSQNIINP--TLIEDIILADRMARATVQELAQRGVSPVA 412

RESULT 5
 Q8SP64 PRELIMINARY; PRT; 472 AA.

AC Q8SP64;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.
 GN X1.
 OS Artemisia annua (Sweet wormwood).
 OC Caryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC eumatiophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC steridae; campanulids; Asterales; Asteraceae; Asteroideae;
 OC Anthemideae; Artemisia.
 OX NCBI_TaxId=35608;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Yugoslavica; TISSUE=Root;
 RA Debe K.K., Scout F.F., Shore K.A., Weathers P.J.;
 RT Artemisia annua 1-deoxy-D-xyulose-5-phosphate reductoisomerase (dxr)
 RT XNA.
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR XBL; AF182287; AAD56391.2;
 DR InterPro; IPR003821; DXP reductoisomase.
 DR Fam; PF02670; DXP reductoisom; 1.
 DR GSRFams; TIGR00243; Dxr; 1.
 KM Isomerase.
 SQ SEQUENCE 472 AA; 50740 MW; D52023C09D475675 CRC64;

Query Match 42.2%; Score 838.5; DB 10; Length 472;
 Best Local Similarity 46.2%; Pred. No. 3.9e-54;
 Matches 181; Conservative 66; Mismatches 134; Indels 11; Gaps 5;

QY 2 KGICILGATGSGVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWMSK 61
 DB 75 KPIIITSGTSGTGYLDIVAGYDRPRVGLAAGNLERIPQIRQFQPIVSIADPE 134
 QY 62 VAEFRQRIASPVDADIKVLSGSEALQOVATLENVDYMAAIVGAAGLLPTLAARAGKT 121
 DB 135 VAEIKBALAGSDVMP-EIIPDEGVVVAARHPDCVTVVTGIVGAGLKPVAALIEAGNVI 193
 QY 122 LLAANKKALVMSGQIFMQAVSDSGAVLPIIDSEHNAIQCM---PAGYTPGHTAQARI 181
 DB 194 ALANKETLIAGGPVTLPIAHKRNKILPADESHSAIFQCLQGVPEG-----ALRKIIIT 248
 QY 182 SGCFPRRTPIETLSSTVDPQAVAHPKMDGKRIYSDATMNNKGLIELIACLLFNM 241
 DB 249 SGCAFDPMPVPELQAVTADALKHPNMSGKPIITVDSATLNKGLIEVIAHYLFQSSYDN 308
 QY 242 IEVVIHPQSIHSMVDYDGSVLAQMGNDPMTPIAHAMAMPEDSDVAALDIFE 298
 DB 309 IDIVIHPOSIIHSMVDYDGSVLAQMGNDPMTPIAHAMAMPEDSDVAALDIFE 368
 QY 299 GGHMDPEKEDLRFPCRLAIVBAIKSGIMPTVLANAEIAVEATLNEVEKFTDIAVIER 358
 DB 369 GSLTFKAPDNVYKYSMSHLAYSAGAGGTGCVLSAANKKAVEMTLDKIGTLDIFKVEL 428
 QY 359 SMAQFKPD--DAGSLBLVLAQADODAREVARDI 388
 DB 429 TCEHQQLVLPASLIERIHYDLWARREYASV 460

RESULT 6
 Q8E6C PRELIMINARY; PRT; 473 AA.

AC Q8E6C;
 DT 1-OCT-2002 (TREMBlrel. 22, Created)
 DT 1-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 1-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.
 GN X1.
 OS Artemisia annua (Sweet wormwood).
 OC Caryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC eumatiophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC steridae; campanulids; Asterales; Asteraceae; Asteroideae;
 OC Anthemideae; Artemisia.
 OX NCBI_TaxId=35608;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Yugoslavica; TISSUE=Root;
 RA Debe K.K., Scout F.F., Shore K.A., Weathers P.J.;
 RT Artemisia annua 1-deoxy-D-xyulose-5-phosphate reductoisomerase (dxr)
 RT XNA.
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR XBL; AF182287; AAD56391.2;
 DR InterPro; IPR003821; DXP reductoisomase.
 DR Fam; PF02670; DXP reductoisom; 1.
 DR GSRFams; TIGR00243; Dxr; 1.
 KM Isomerase.
 SQ SEQUENCE 473 AA; 50740 MW; D52023C09D475675 CRC64;

RT "Cloning and expression of cDNAs encoding two enzymes of the ME
RT pathway in *Stevia rebaudiana* Bertoni.",
RL Plant Physiol. 0:0-0(0).
DR EMBL: A0729233; CAD22156.1; -.
DR InterPro: IPRO03821; DXP_redoxism.
DR Pfam: PF02670; DXP_redoxism; 1.
DR TIGRfam5: TIGR00243; Dxr; 1.
KW Isomerase.
SQ SEQUENCE 473 AA; 51021 MW; 03BD5BP204B62490 CR664;

Query Match	41.7%;	Score 827.5;	DB 10;	Length 473;
Best Local Similarity	45.4%;	Pred. No. 2.6e-53;		
Matches 179;	Conservative 71;	Mismatches 129;	Indels 15;	Gaps 6

Qy	2	KGICIGAGSGVGVTLVVAHAPKXVVALTANGIDALYECCLAHPEYAVVMSEK	61
Dp	76	KPISIVGSGISICTQTLDIYAENPKFRVVALAAGSVITLARGIRAFKQVLSIONBL	135
Qy	62	VAEFKRIAAFPVADI--KVLSGSEALQQVATLENVDVTMAAIVGAAGLPTLAANKK	119
Dp	136	VGEIYKELALD--ADYMEPIIFGQDGLIEVARHDPCTVVTGIVGCGILPTVAALIAEK	192
Qy	120	TYLANKELVWSGQITPMAVSQSAVLLPTDSEHNAIFQMAGTTPGHTAKQAPRILL	179
Dp	193	NIALAKRETLIAGGFVLPARKINVKLIPADSEHSAIFQCI-QGPEG---ALRILL	247
Qy	180	TASGGEFRTPPIETLSSVTPDQAAVHPMDGRKISVDSATMMKIGLEILEACLFENEP	239
Dp	248	TASGAFRLPPEBKLDYKAVADALGHPMSMGKKITVDSATLRKGLIYEATLYGSDY	307
Qy	240	DOIEVVIHQSIHHSNVDYDGSVLAQWMDMPTIAHAMAPERDSCVAD--LDIF	296
Dp	308	DNEIIVIHQSIHHSNVDYDGSVLAQWMDMLPTLITLSMPDRISCSIEITMPEILDIC	367
Qy	297	EYGHNDPEKPDKRPPLRLAYEAAKSGGIPPTYLANANELVAPLANEYKTTDIAVIL	356
Dp	368	KIGSLIFKAPDWKYPSPNDLAAARAGSGITGVLASANEKAVEMFIDEKITYLDIFKVV	427
Qy	357	ERSMQFKPD--DAGSLLEYQADDDASEVARDI	388
Dp	428	ELTCAKGSSELVAPLSLEIYHYIDMARDYASL	461

RESULT 7		
08D2G6		
ID	Q8D2G6	PRELIMINARY; PRF; 397 AA.
AC	Q8D2G6;	
DT	01-MAR-2003 (TRIMBLrel. 23, Created)	
DT	01-MAR-2003 (TRIMBLrel. 23, Last sequence update)	
DT	01-MAR-2003 (TRIMBLrel. 23, Last annotation update)	
DE	YaeM protein.	
GN	YAE.M.	
OS	Miglesworthia brevipalpis.	
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
OC	Enterobacteriaceae; Miglesworthia.	
OX	NCBI_TaxId=164609;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	MEDLINE=22297718; PubMed=12219091;	
RA	Akman E., Yamashita A., Matanabe H., Oshima K., Shida T., Hattori M.	
RA	Aksoy S.;	
RT	"Genome sequence of the endocellular obligate symbiont of tsetse	
RT	fly files, Miglesworthia glossinidia."	
RL	Nat. Genet. 32:402-407(2002).	
DR	EMBL: AB063522; BAC24534.1; -	
DR	Complete proteome.	
QC	SEQUENCE 397 AA; 44663 MW; 485355F5256BD56 CRC64;	

Query Match	41.2%;	Score 818.5;	DB 16;	Length 397;
Best Local Similarity	42.4%;	Pred. No. 9,4e-53;		
Matches 171;	Conservative 81;	Mismatches 130;	Indels 21;	Gaps 4
Qy	1	MKGICLGATGSGUSTLDVARHPDKQVWALTANG-NIDALVEGCLAHHPREAVVME	59	

Db 1 MKKITLITSSGSI GNKITLKITSNMLDKRSVSLVAYGNINIVLLISQCIKRYPNVYCIENK 60

Qy 60 SKVAEFGKSIASPVADIKVLSGSEALQOVATLENVDTYMAALYGAAGLEPTLAAAGK 119

Db 61 KKLIDTKQGLKNK -CKTSVLPFGSNDICMLSSSKSEVDIVISATVLSGIFLFPFAISGK 119

Qy 120 TYLLAKKALVMSGQIFMQAVSDSGAVLLPDISEHNAIFQCPMAGYTPGHTAKAR --- 175

Db 120 KILANKEILVSCGHPFKQVEXKMSIILPIDSEHNAIFQSLPLDF -----QKLGIA 172

Qy 176 -----RILITASGSPRRTPPIETLSSVTPDQAVAHPKMDGRKI SVDSATYMNKGL 22

Db 173 SIINKYGIYKLITLTSGSGPFRNWEFLDKVSPDQACHPMKMKKSIIDSATYMNKGF 223

Qy 228 LIEACLLFMNEPDQIEVVTIHPOSIIHSMVWDVSDSLAQMGNPMRTPIAHMAMPERP 287

Db 223 YIYAKULFVACDQIEELIHHOSIYHSMIRKIIDGIVLANSLEPMOSSISGLOPRK 297

Qy 288 SGVAPLDIFEVGMDFEKPDLEKRFQCLAYEALIKSGIMEFTVMAANEIYAEFLNEEV 347

Db 293 IKKKYLDYFKNKKLTFPESIDYKRYECLIMLALQASNGAGATTVINSANEISVSFSLSKI 352

Qy 348 KPTDIAVITERSMAQFKPDGSLGVLYQADQARAYARDIK 390

Db 353 YFPDIAIINKVLDKIDLFEPSSIEELIILDSKARNTAKKETIK 395

```

RESTIT 8
ID 09473 PRELIMINARY: PRT: 475 AA.
AC 0947C3
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, last annotation update)
DE 1-deoxy-D-xylinulose-5-phosphate reductoisomerase.
GN DXR.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21425086; PubMed=11532167;
RA Rodriguez-Concepcion M., Ahumada I., Diez-Juez E., Saurer-Gueto S.,
RA Itois L.M., Gallego F., Carretero-Paulet L., Campos N., Boronat A.;
RT 1-deoxy-D-xylinulose 5-phosphate reductoisomerase and plastid
RT leptonoid biosynthesis during tomato fruit ripening. ";
RL Plant J. 27:213-222(2001).
DR EMBL; AF31705; AK95063.2; -.
DR InterPro; IPR003821; DXP reductoisom.
DR Pfam; PF02670; DXP reductoisom; 1.
DR TrIRPams; TIGR00243; Dxr; 1.
DR Isomerase.
SQ SEQUENCE 475 AA; 51497 MW; E17742A3BB0C0782 CRC64;

```

```
QY QY Match          41.1%; Score 816.5; DB 10; Length 475;
Be % Local Similarity 45.1%; Pred No. 1.7e-52;
Ma ches 180; Conservative 69; Mismatches 125; Indels 25; Gaps 7

QY      2 KGICIGATGSGISGVSTLDVVAHPDKGYVALTANGNIDALYEQCLAHPEYAVVMESK 61
Db      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      79 KPISIVGSVSGISGTOTLDAIENPDKFRVALLAGSNVTLLADQYKTRPKLVARNBSL 138
Db      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY      62 VAEFKRIASPVADI-----KVLSSSEKLQQVATLENVDTMALIYGAAGLLPITLAAGA 11
Db      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY     139 VEELKDAL-----AMDEKPELIIPBEOQVIEVARHPDPAVTAVTGIVCGAKIKPTVALEA 19
Db      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY    118 GKTVLLANKREALVMSGOQLFMQAVSDSGAVLLPIDBEHARFQC---MPAGYTPGHATAQA 174
Db      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY   194 GKQIALANKEETLIAGGFPLVPAPAHKKXKIIPADSBHSALPCCIQGLEPG-----AL 244
Db      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

QY 175 REILLTASGPRRTPIETLSVTDPQAAHKKMKERKISVDSATMMKGLIEACLL 234
 DB 246 REILLTASGPRRTPIETLSVTDPQAAHKKMKERKISVDSATMMKGLIEACLL 305
 QY 235 FMEPDDQIEVVIHPOSIHSMVYDGSVLAQMGNDKRTPIAHAMPER-FDSGVA-- 291
 DB 306 FGAEDNIEIVIHPOSIHSMVETQDSSVLAQMGNDKRTPIAHAMPER-FDSGVA-- 365
 QY 292 FMEPDDQIEVVIHPOSIHSMVYDGSVLAQMGNDKRTPIAHAMPER-FDSGVA-- 351
 DB 366 FGAEDNIEIVIHPOSIHSMVETQDSSVLAQMGNDKRTPIAHAMPER-FDSGVA-- 425
 QY 352 IAVIERSAOKRPPD--DAGSELVLAQMGNDKRTPIAHAMPER-FDSGVA-- 388
 DB 426 IPRIVELTCAHREELVSSPSLEIITHYDMARDYAAST 464

RESULT 9

Q8W250 PRELIMINARY; PRT; 473 AA.
 AC 08W250;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase precursor
 (EC 1.1.1.-).
 OS Oryza sativa (Rice).
 OC Burkholderia: Vitridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Burharoidae; Oryzaceae; Oryza.
 NCBI_TaxID=4530;
 (1)
 RA SEQUENCE FROM N.A.
 RP Caretero-Paulet L., Boronai A., Campos N.;
 RT 11-deoxy-D-xyulose 5-phosphate reductoisomerase (DXR), catalyzing the
 first committed step of the mevalonate-independent pathway for IPP
 biosynthesis.
 RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF367205; AAL37560.1; -.
 DR Gramene; Q8W250; -.
 DR InterPro; IPR003821; DXR_reductoisomase.
 DR Pfam; PF02670; DXR_reductoisom; 1.
 DR TIGRfam; TIGR00243; Dxr; 1.
 KW Transit peptide; Isomerase; Oxidoreductase.
 FT TRANSIT 1
 SQ SEQUENCE 473 AA; 51497 MW; 618C4A675283478F CRC64;

Query Match 40.8%; Score 810.5; DB 10; Length 473;
 Best Local Similarity 45.4%; Pred. No. 4.8e-52;
 Matches 182; Conservative 67; Mismatches 123; Indels 29; Gaps 8;

QY 2 KGCITLIGATGSGVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHHPYAVVWESK 61
 DB 77 KPISIVGSGTIGTQTLDIVAENPKFRVVALAAGSNVTLADQVTKFKPKVAVRNBSL 136
 QY 62 VAEFRQRIASPVADIK--VLSGSEALQOVATLENDVTVMAIYGAAGLPTLAAARAG 119
 DB 137 VDELKEALAD--CDMKPEIIPGEGVTEVAARHPDVAITVGTIGAGLKPVAALTEAGK 193
 QY 120 TYLANKEALVMSGOIFMOAVSDSGAVLLPDSSENAIFOC--MPAGTTPGHTAKQARR 176
 DB 194 DIALANKETLIGGPFVPLAOKHKVILIPADSEHSAIFOCIGLPEG-----ALRR 245
 QY 177 ILLTASGPRRTPIETLSVTDPQAAHKKMKERKISVDSATMMKGLIEACLLFN 236
 DB 246 ILLTASGPRRTPIETLSVTDPQAAHKKMKERKISVDSATMMKGLIEACLLFN 305
 QY 237 MEPPDQIEVVIHPOSIHSMVYDGSVLAQMGNDKRTPIAHAMPER-FDSGVA--PL 293
 DB 306 AEDYDIEIVIHPOSIHSMVETQDSSVLAQMGNDKRTPIAHAMPER-FDSGVA--PL 365
 QY 294 DIFEVGNDKPRKPPCLALAYEAIKSGIMPTVLANAEIAYAPFNEBYKFTDIA 353

DB 366 DLCKGSLTEKAPNVKYPSPMDLAAAGRAGTMTGVLASNAEKAVHFLIDEKIGYDIF 425
 QY 354 VITE-----RSMQKRPDDAGSLEVLQADQDAREVARDI 388
 DB 426 KVELTCDARHNELVTSR-----SLEIITHYDMAREVAAST 464

RESULT 10

Q9FTX PRELIMINARY; PRT; 473 AA.
 AC 9FTX;
 DT 1-MAR-2001 (TREMBLrel. 16, Created)
 DT 1-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 1-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.
 GN 0005A05.19 OR P0482C06.2.
 OS Oryza sativa (Rice).
 OC Burkholderia: Vitridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Burharoidae; Oryzaceae; Oryza.
 NCBI_TaxID=4530;
 (1)
 RA SEQUENCE FROM N.A.
 RP TRAIN=cv. Nipponbare;
 RC TRAIN=cv. Nipponbare;
 RA Asaki T., Matsumoto T., Yamamoto K.;
 RT Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0005A05.19;
 RT Submitted (SRP-2000) to the EMBL/GenBank/DBJ databases.
 (2)
 RA SEQUENCE FROM N.A.
 RP TRAIN=cv. Nipponbare;
 RC TRAIN=cv. Nipponbare;
 RA Asaki T., Matsumoto T., Yamamoto K.;
 RT Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0482C06.2;
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 (3)
 RA Asaki T., Matsumoto T., Yamamoto K.;
 RT Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0482C06.2;
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

Que / Match 40.7%; Score 807.5; DB 10; Length 473;
 Best Local Similarity 45.4%; Pred. No. 8e-52;
 Matches 182; Conservative 66; Mismatches 124; Indels 29; Gaps 8;

QY 2 KGCITLIGATGSGVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHHPYAVVWESK 61
 DB 77 KPISIVGSGTIGTQTLDIVAENPKFRVVALAAGSNVTLADQVTKFKPKVAVRNBSL 136
 QY 62 VAEFRQRIASPVADIK--VLSGSEALQOVATLENDVTVMAIYGAAGLPTLAAARAG 119
 DB 137 VDELKEALAD--CDMKPEIIPGEGVTEVAARHPDVAITVGTIGAGLKPVAALTEAGK 193
 QY 120 TYLANKEALVMSGOIFMOAVSDSGAVLLPDSSENAIFOC--MPAGTTPGHTAKQARR 176
 DB 194 DIALANKETLIGGPFVPLAOKHKVILIPADSEHSAIFOCIGLPEG-----ALRR 245
 QY 177 ILLTASGPRRTPIETLSVTDPQAAHKKMKERKISVDSATMMKGLIEACLLFN 236
 DB 246 ILLTASGPRRTPIETLSVTDPQAAHKKMKERKISVDSATMMKGLIEACLLFN 305
 QY 237 MEPPDQIEVVIHPOSIHSMVYDGSVLAQMGNDKRTPIAHAMPER-FDSGVA--PL 293
 DB 306 AEDYDIEIVIHPOSIHSMVETQDSSVLAQMGNDKRTPIAHAMPER-FDSGVA--PL 365
 QY 294 DIFEVGNDKPRKPPCLALAYEAIKSGIMPTVLANAEIAYAPFNEBYKFTDIA 353
 DB 366 DLCKGSLTEKAPNVKYPSPMDLAAAGRAGTMTGVLASNAEKAVHFLIDEKIGYDIF 425
 QY 354 VITE-----RSMQKRPDDAGSLEVLQADQDAREVARDI 388

DB 426 KVELTCDARHNEVTRP---SLEETIHYDLMAREYASL 462

RESULT 11

OSM4M4 PRELIMINARY; PRT; 474 AA.

AC OSM4M4
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE 1-deoxy-D-xylulose-5-phosphate reductoisomerase.
 DXR.
 OS Catharanthus roseus (Rose periwinkle) (Madagascar periwinkle).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Gentianales; Apocynaceae; Rauvolfiaceae; Vinaceae;
 OC Catharanthus.
 OC NCBI_Taxid=4058;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20568781; PubMed=11118631;
 RA Veau B., Courtois W., Oudin A., Chenieux J.-C., Rideau X., Claeste M.;
 RT "Cloning and expression of cDNAs encoding two enzymes of the MEP
 pathway in Catharanthus roseus."
 RL Biochim. Biophys. Acta 1517:159-163 (2000).
 DR EMBL; AF250235; AAF65154.1; -
 DR InterPro: IPR003821; DXP_reductoisomase.
 DR Pfam: PF02670; DXP_reductoisom; 1.
 DR TIGRFAMs; TIGR00243; Dxr; 1.
 KM Isomerase.
 SQ SEQUENCE 474 AA; 51165 MW; C2737CD318B0C94C CRC64;

Query Match 40.5%; Score 803.5; DB 10; Length 474;
 Best Local Similarity 44.0%; Pred. No. 1.6e-51;
 Matches 176; Conservative 71; Mismatches 134; Indels 19; Gaps 6;

QY 2 KGIICLGATSGISGVSTLDVVAARHPDKYQVVALTANENIDALVEQCLAHPEYAVVWESK 61
 DB 78 KPISIVSGTSGVGTCTLDIVAENPKFRVVALAGSNVTLADQVTKFKPELVAVRNESL 137
 QY 62 VAEFKORLASPVAD-IKLGSSEALQVATLENDVTMAALVGAAGLPTLAAGKT 120
 DB 138 VNSLKE--ALSDVDDPEEITPGQGVVEVRHSDAVTVGTGVGAGLKPVALAELAGKD 195
 QY 121 VILANKKALVMSGOIFMQAVSDSGAVLLPIDSEHNAIFQCC--MDAGYTPGHTAQARRI 177
 DB 196 IALANKETILACKPVLPLAHKHKVILPDADEHSAIFQCIQGLPEG-----ALRRI 247
 QY 178 LITASGGPRRTPIETLSSTVPDQAVAHPRKMDGKISVDSATMMNKGILIEACLITRM 237
 DB 248 ILTASGGAARDMPVEELKAVKADALKHPNMWNGKKTIVDSATLTFNKGLEVEIAHYLFQA 307
 QY 238 EPPDIEVVIHPOSIHSMVDYDGSVLAQMGNDPKRTPIAHAMWPERFDSGVAP--LD 294
 DB 308 EVDNDIVIHPOSIHSMVETODSSVLAQLGMPDMRLPLITLTSMPDRISCEIIMPRLD 367
 QY 295 IREVGMDPEKPDLPKFPCLRLAYBAIKSGGIMPTVLANAEIAYEAPINEVEKTTDLAV 354
 DB 368 LCKLGLSLFTKTPDNVYKPSMDLAVAAAGRAGGTGTGTLASANKKAVELPIDEKISYLDIFK 427
 QY 355 IERSMAOKPD--DAGSLELVLAQDQDARVARDIKTL 392
 DB 428 VVELTGNARHNELVTPSLDEIIVHYDLGARDYAASTQNSL 467

RESULT 12

OSM4M4 PRELIMINARY; PRT; 472 AA.

AC OSM4M4
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Putative 1-deoxy-D-xylulose 5-phosphate reductoisomerase.
 CN DXR.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
 OC NCBI_Taxid=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=dwaft-1; TISSUE=Root;
 RA Hans J., Haube B., Strack D., Walter M.H.;
 RT "Regulation of the non-mevalonate methylerythritol phosphate (MEP)
 pathway by mycorrhizal fungi."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ297566; CAC03581.1; -
 DR InterPro: IPR003821; DXP_reductoisomase.
 DR Pfam: PF02670; DXP_reductoisom; 1.
 DR TIGRFAMs; TIGR00243; Dxr; 1.
 KM Isomerase.
 SQ SEQUENCE 472 AA; 51285 MW; 9D5E30F8C8A0706C CRC64;

Query Match 40.4%; Score 802.5; DB 10; Length 472;
 Best Local Similarity 45.1%; Pred. No. 1.9e-51;
 Matches 178; Conservative 68; Mismatches 132; Indels 17; Gaps 6;

QY 2 KGIICLGATSGISGVSTLDVVAARHPDKYQVVALTANENIDALVEQCLAHPEYAVVWESK 61
 DB 76 KPISIVSGTSGVGTCTLDIVAENPKFRVVALAGSNVTLADQVTKFKPELVAVRNESL 135
 QY 62 VAEFKORLASPVAD-IKLGSSEALQVATLENDVTMAALVGAAGLPTLAAGKT 118
 DB 136 VDELKRALACERKP---ETIPGQGVIEVAHPDVAVTGTGVGAGLKPVALAELAG 191
 QY 119 KTVILANKKALVMSGOIFMQAVSDSGAVLLPIDSEHNAIFQCCMDAGYTPGHTAQARRI 178
 DB 192 KDILANKETILACKPVLPLAHKHKVILPDADEHSAIFQCI---QGLSEALRRI 246
 QY 179 LITASGGPRRTPIETLSSTVPDQAVAHPRKMDGKISVDSATMMNKGILIEACLITRM 238
 DB 247 ILTASGGAARDMPVEELKAVKADALKHPNMWNGKKTIVDSATLTFNKGLEVEIAHYLFQA 306
 QY 239 EPPDIEVVIHPOSIHSMVDYDGSVLAQMGNDPKRTPIAHAMWPERFDSGVAP--PLDI 295
 DB 307 YDDIEIYIHPOSIHSVETODSSVLAQLGMPDMRLPLITLTSMPDRISCEIIMPRLD 366
 QY 296 FEVGMDPEKPDLPKFPCLRLAYBAIKSGGIMPTVLANAEIAYEAPINEVEKTTDLAVI 355
 DB 367 CKLGSITFRAPDNVYKPSMDLAVAAAGRAGGTGTGTLASANKKAVELPIDEKISYLDIFK 426
 QY 356 IERSMAOKPD--DAGSLELVLAQDQDARVARDIKTL 388
 DB 427 VVELTGNARHNELVTPSLDEIIVHYDLGARDYAASTQNSL 461

RESULT 13

OSM4M4 PRELIMINARY; PRT; 391 AA.

AC OSM4M4
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE 1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.-).
 CN DXR OR LA3292.
 OS Leptosiphia interogens.
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
 OC NCBI_Taxid=173;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;
 RA Ren S.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF011490; AAN50490.1; -
 KM Isomerase; Oxidoreductase; Complete proteome.

```

SQ SEQUENCE 391 AA; 43026 MW; 8CF56BC0C17AC0F7 CRC64;
Query Match 37.1%; Score 735.5; DB 16; Length 391;
Best Local Similarity 41.0%; Pred. No. 1,4e-46;
Matches 158; Conservative 78; Mismatches 140; Indels 9; Gaps 4;
QY 4 ICILGATGSGVSTLDVVAHPDK---YOVVALTANANGIDALYEGCLAHPEYAVVME 63
DB 7 VCLLDAGSGVSGSTLVAHPDKRLHSFVSHLSKAKEIQKESFDP--ICVSSDFA 64
QY 64 BEKORIASPVADIKVLSGSEALQOVATLENDVYMAAIVGAAGLLPTLAAKAKGTVLL 123
DB 65 DV--GVLANKLGRTQLVIGSESLCELVEPEVEIYITAVGSGVGLRPTIAITGKRLGI 122
QY 124 ANKELAVMSGQIFMQAVSDGAVLPIFSEHNATIQCPAGTPEHTKQARRILLTMSG 183
DB 123 ANKELAVMSGQIFMQAVSDGAVLPIFSEHNATIQCPAGTPEHTKQARRILLTMSG 177
QY 184 GFRRTPIETLSSVTPDOVAHPKMDGRKISVDSATMNNKLEIEACLEFNNEPDQIB 243
DB 178 GAFRLPVEFGJSSVTKEGALHPTMNGKPTITIDNGKINGLEVIEHLEFNPYDKIG 237
QY 244 VVIHPQSTIHSNVDYDGSVLAQMGNDPMKTPPIAHAMAPERPDSGVALDIFVEGMDP 303
DB 228 VVIHPQSTIHSNVDYDGSVLAQMGNDPMKTPPIAHAMAPERPDSGVALDIFVEGMDP 297
QY 304 EKPDLKRPCLRLAYEARLKSQGMPTVLANEIAVEAFNEHVEFTDIAVIERSMAGP 363
DB 238 REPDKRIPGLAGLAEGAGVGATPCITFNAANEAVALFLKOEIRFIEIPYIRETDEI 357
QY 364 KPDDAGSLELVLAQADQAREVARDI 368
DB 358 KIEFPLSLBEYEAEDRIARETVRN 382
RESULT 14
ID 096693 PRELIMINARY; PRT; 488 AA.
AC 096693;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2003 (TREMblrel. 21, Last annotation update)
DN 1-deoxy-D-xyulose 5-phosphate reductoisomerase.
GN DXR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxId=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB3;
RA Altincioek B., Sanderbrand S., Wiesner J., Jomaa H.;
RT "dxr as a potential target for antimalarial drugs.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF111813; AAD03739.1; -
DR InterPro: IPR003821; DXR_reductoisom.
DR Pfam: PF02670; DXR_reductoisom; 1.
DR Trifams; TRG00243; DXR; 1.
KW isomerase.
SQ SEQUENCE 488 AA; 55756 MW; 4E280C81CDFAD3EF CRC64;
Query Match 36.8%; Score 729.5; DB 5; Length 488;
Best Local Similarity 37.8%; Pred. No. 5,4e-46;
Matches 153; Conservative 88; Mismatches 141; Indels 23; Gaps 7;
QY 4 ICILGATGSGVSTLDVVAHPDK---YOVVALTANANGIDALYEGCLAHPEYAVVME 59
DB 80 VAIFGSGSIGTALNII-RECNKIEVFNKALYVKSVMLEYEQAREPLPEYLCHDK 138
QY 60 SKVAEPFORIASPVADIK--VLSGSEALQOVATLENDVYMAAIVGAAGLLPTLAAKA 117
DB 139 SVTEELKEV--KNIKDYKPIILCGDEGMEICSSNSIDKIVIGDSFGQGLSTMYATMN 196
QY 118 KTVLLANKKALVMSGQIFMQAVS--DSGAVLLPIDESEHNAIFQCPAGTPEHTAKQ--- 173

```

```

DB 197 NCVVLANKESIVSAGFPLKLLNIHKNAKTIPVDSHSAIPQCLDNKKVLTCKLQDNF 256
QY 174 -----AARILLTASGFFRRPTIETLSSVTPDOVAHPKMDGRKISVDSATMNNKLEI 228
DB 257 SKINNINKIFLCSSGGFFQNLTWDELKNVTSNALKHPKMGKIKTIDATMNNKLEI 316
QY 229 IEACLFNNEPDQIEVVIHQSTIHSNVDYDGSVLAQMGNDPMKTPPIAHAMAPERFDS 288
DB 317 IETHFLPDVVDNIIEVIVHKECIHSCEPIDSVSQMYPPMQPIPLVSLTPMDIKT 376
QY 289 GVALPDIIEFGHNDPEKPPCLRLAYEARLKSQGMPTVLANEIAVEAFNEHVEK 348
DB 377 NLRPLDAQVSTLTFHPSLEHPPCLRLAYQAGIKNFYTVLANSEIANNLEPANNKIK 436
QY 349 FTDIAVIERSMAGKPDAGS-----LELVLAQADQAREVARDI 388
DB 437 YFDISIIISQVLESFNSQKSENSIDLKQILQIHSYAKKATDI 481
RESULT 15
ID 091K4 PRELIMINARY; PRT; 488 AA.
AC 091K4;
DT 1-MAR-2003 (TREMblrel. 23, Created)
DT 1-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 1-MAR-2003 (TREMblrel. 23, Last annotation update)
DN 1-deoxy-D-xyulose 5-phosphate reductoisomerase.
GN 114_0641.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxId=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA 3DLINE=22255705; PubMed=12368664;
RA Irwin M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,
RA Wilson J.M., Pain A., Nelson K.B., Bowman S., Paulsen I.T., James K.,
RA Jensen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA van M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Anguilo S.,
RA Irwin D.M.A., Fairhead A.H., Fraunholz M.J., Roos D.S., Ralph A.B.,
RA Madden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Hunter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Mascher C.M., Barrrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum";
RL Nature 419:498-511(2002).
DR EMBL; AB014826; AAN37254.1; -
KW isomerase.
SQ SEQUENCE 488 AA; 55756 MW; 4E280C81CDFAD3EF CRC64;
Query Match 36.8%; Score 729.5; DB 5; Length 488;
Best Local Similarity 37.8%; Pred. No. 5,4e-46;
Matches 153; Conservative 88; Mismatches 141; Indels 23; Gaps 7;
QY 4 ICILGATGSGVSTLDVVAHPDK---YOVVALTANANGIDALYEGCLAHPEYAVVME 59
DB 80 VAIFGSGSIGTALNII-RECNKIEVFNKALYVKSVMLEYEQAREPLPEYLCHDK 138
QY 60 SKVAEPFORIASPVADIK--VLSGSEALQOVATLENDVYMAAIVGAAGLLPTLAAKA 117
DB 139 SVTEELKEV--KNIKDYKPIILCGDEGMEICSSNSIDKIVIGDSFGQGLSTMYATMN 196
QY 118 KTVLLANKKALVMSGQIFMQAVS--DSGAVLLPIDESEHNAIFQCPAGTPEHTAKQ--- 173
DB 197 NCVVLANKESIVSAGFPLKLLNIHKNAKTIPVDSHSAIPQCLDNKKVLTCKLQDNF 256
QY 229 IEACLFNNEPDQIEVVIHQSTIHSNVDYDGSVLAQMGNDPMKTPPIAHAMAPERFDS 288
DB 317 IETHFLPDVVDNIIEVIVHKECIHSCEPIDSVSQMYPPMQPIPLVSLTPMDIKT 376
QY 289 GVALPDIIEFGHNDPEKPPCLRLAYEARLKSQGMPTVLANEIAVEAFNEHVEK 348
DB 377 NLRPLDAQVSTLTFHPSLEHPPCLRLAYQAGIKNFYTVLANSEIANNLEPANNKIK 436
QY 349 FTDIAVIERSMAGKPDAGS-----LELVLAQADQAREVARDI 388
DB 437 YFDISIIISQVLESFNSQKSENSIDLKQILQIHSYAKKATDI 481

```

```

Db      317  IETHPLFDVNDYNDIEVIVHKECIIHSCVEFIDKSVISQMYYTDMOQIPLISLTWPDRIKE 376
Qy      289  GVAPLIDIEFVGHMDPEKPDILKRFPCILRLAYEAIKSGGIMPTVLNANEIAYEAPINEEVK 348
Db      377  NLKPLDLAQVSTLTFPHKPSLEHFPCTIKLAYQAGIKGNFYPTVLNANSNEIANNLFINNKIK 436
Qy      349  FTDIAVITIERSMQKFPDDAGS-----LEVLQADQDAREVARDI 388
Db      437  YFDISSIISQVLESEFNSQKVSSENSDLMKQIILQIHSWAKDKATDI 481

```

Search completed: January 29, 2004, 15:54:04
 Job time : 31.3684 secs

APPLICANT: Wang, Min
 APPLICANT: Mooney, Jeffrey L.
 APPLICANT: Debouck, Christine M.
 APPLICANT: Zhong, Yi Yi
 TITLE OF INVENTION: No. 6204042e1 GimU
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESS: Dechert Price & Rhoads
 STREET: 4000 Bell Atlantic Tower, 1717 Arch Street
 City: Philadelphia
 STATE: PA
 COUNTRY: US
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: PASTESQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/309,026
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/971,782
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Dickinson, Todd Q
 REGISTRATION NUMBER: 28,354
 REFERENCE/DOCKET NUMBER: GM10024
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-994-2252
 TELEFAX: 215-994-2222
 TELEX:
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 459 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 /S-09-309-026-2

Query Match	7.4%;	Score 88.5;	DB 3;	Length 459;
Best Local Similarity	23.9%;	Pred. No. 0.12;		
Matches	48;	Conservative	32;	Mismatches 74;
				Indels 47;
				Gaps 9;

```

Qy      8 MAVVPAAGCKRQAOAPQCVPLKAGTITHTTLRLLESFAFOKVAIVSEVPWPEL 67
Db      4 FALILAGGCTRKSKSLPYLHKVAGISMLH-----VFSVC-AI0PE----- 46

Qy      68 SIARHPDIIITAPGKERADSVLSALKALEDIASENDVYLVDHA--ARPCLTGSDIHLQI 124
Db      47 -----KTVLVGHR--AEIVAEVLAGQEPETQSEOGTGHAAVMTEPILLEGISGHTLV 98

Qy      125 DTLKNDPVGGILALSHDTLKHVGGDTTATITDRKGVRAALTPOM--FKYGM-----R 176
Db      99 -----IAGDPPLITGSESLKNL-----IDFHNRKVAITILTAEIDNPFGRIVRADNA 147

Qy      177 DALQRTGNGPAVTDASALEL 197
Db      148 EVLRMVEQDADDFEKQIKETI 168

```

Search completed: January 29, 2004, 15:57:04
Job time : 8.2499 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 ; Search time 15.7685 Seconds
(without alignments)
3044.503 Million cell updates/sec

Title: US-09-941-947a-10

Perfect score: 1199
Sequence: 1 MNPPTICMAVVPAGVGKRM.....IKITRPEDLALQPTMEQA 231

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEM_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEM_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09A_NEM_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1199	100.0	231	10 US-09-934-903-6	Sequence 68, Appli
2	1199	100.0	231	10 US-09-934-868-6	Sequence 10, Appli
3	1199	100.0	231	11 US-09-941-947a-10	Sequence 11931, A
4	555.5	46.3	234	9 US-09-815-242-11931	Sequence 11076, A
5	555.5	44.2	225	9 US-09-815-242-11076	Sequence 11072, A
6	529.5	44.2	254	9 US-09-815-242-11722	Sequence 10273, A
7	520.5	43.4	236	9 US-09-815-242-10273	Sequence 11, Appli
8	493.5	41.7	236	9 US-09-792-251-11	Sequence 13667, A
9	493.5	41.7	236	9 US-09-815-242-13667	Sequence 8, Appli
10	283.5	24.8	232	10 US-09-792-251-8	Sequence 8, Appli
11	283.5	23.6	302	10 US-09-935-943-8	Sequence 6, Appli
12	255.5	21.3	232	12 US-10-128-713A-6	Sequence 11503, A
13	245.5	20.5	250	15 US-10-156-761-11503	Sequence 621, App
14	240.5	20.1	218	12 US-10-289-762-621	Sequence 6422, Ap
15	223.5	18.6	256	10 US-09-738-626-6422	

1	217.5	18.1	228	9 US-09-815-242-12471	Sequence 12471, A
1	214.5	17.9	227	9 US-09-815-242-5711	Sequence 5711, Ap
1	214.5	17.9	238	9 US-09-816-181-2	Sequence 2, Appli
1	209	17.4	235	9 US-09-815-242-13256	Sequence 13256, A
2	209	17.4	235	10 US-09-792-251-2	Sequence 2, Appli
2	198.5	16.6	235	10 US-09-816-182-2	Sequence 5707, Ap
2	193.5	16.1	236	9 US-09-815-242-5707	Sequence 12467, A
2	193.5	16.1	242	9 US-09-815-242-12467	Sequence 658, App
2	126	10.5	125	9 US-09-867-550-658	Sequence 21638, A
2	105.5	8.8	291	12 US-10-369-493-21638	Sequence 10937, A
2	98.5	8.2	461	9 US-09-815-242-10937	Sequence 5017, Ap
2	98.5	8.2	783	10 US-09-738-626-5017	Sequence 18528, A
2	98	8.2	395	12 US-10-369-493-18528	Sequence 20306, A
2	94	7.8	327	12 US-10-369-493-20306	Sequence 17446, A
3	93	7.8	297	12 US-10-369-493-17446	Sequence 245, App
3	93	7.8	357	12 US-10-312-273-245	Sequence 487, App
3	93	7.8	357	12 US-10-289-762-487	Sequence 21686, A
3	91	7.6	419	12 US-10-369-493-21686	Sequence 1153, Ap
3	90.5	7.5	285	12 US-10-369-493-1153	Sequence 10400, A
3	90	7.5	456	9 US-09-815-242-10400	Sequence 264, App
3	89.5	7.5	381	16 US-10-080-170-264	Sequence 21335, A
3	89.5	7.5	385	12 US-10-369-493-21335	Sequence 528, App
3	89	7.4	154	9 US-09-939-980-528	Sequence 8238, Ap
3	89	7.4	704	15 US-10-128-714-8238	Sequence 23443, A
4	89	7.4	1317	12 US-10-369-493-1317	Sequence 4551, Ap
4	88.5	7.4	361	12 US-09-738-626-4551	Sequence 11824, A
4	88.5	7.4	485	10 US-10-369-493-11824	Sequence 14528, A
4	88	7.3	394	12 US-10-369-493-14528	Sequence 14505, A
4	87.5	7.3	443	15 US-10-156-761-14505	
4	87.5	7.3	831	15 US-10-156-761-14505	

ALIGNMENTS

RESULT 1	US-09-934-903-6	Application US/09934903
Sequence 6	US20020102690A1	
Pat. No.	US20020102690A1	
GENE/2AL INFORMATION:		
APPLICANT:	Kofias, Mathcos	
APPLICANT:	Odum, J. Martin	
APPLICANT:	Schenzle, Andreas J.	
APPLICANT:	No. US20020102690A1ton, Kelley C.	
APPLICANT:	Tomb, Jean-Francois	
APPLICANT:	Rouviere, Pierre	
APPLICANT:	Picatsoglio, Stephen	
TITLE OF INVENTION:	Genes Involved in Isoprenoid Compounds Production	
PI/3 REFERENCE:	CU1646 US NA	
CURRENT APPLICATION NUMBER:	US/09/934,903	
CURRENT FILING DATE:	2001-08-22	
PREVIOUS APPLICATION NUMBER:	60/229,907	
PREVIOUS FILING DATE:	September 1, 2001	
NUMBER OF SEQ ID NOS:	24	
SOFTWARE:	Microsoft Office 97	
SEQ ID NO 6		
LENGTH:	231	
TERT: PRT		
ORGANISM:	Methylomonas 16a	
FEATURE:		
OTHER INFORMATION:	Amino acid sequences encoded by ORF3	
US-09-934-903-6		
Query / Match	100.0%; Score 1199; DB 10; Length 231;	
Best Local Similarity	100.0%; Pred. No. 2.1e-115;	
Matches	231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Cy	1 MNPPTICMAVVPAGVGKRMADRPKYLPLAGCTVTEHTLRLSDAPKVAVALSV	60
Db	1 MNPPTICMAVVPAGVGKRMADRPKYLPLAGCTVTEHTLRLSDAPKVAVALSV	60
Cy	61 DPYPRISIAKHPIITAPGCKERADSVLSALRALIEDIASNDVVLVHDARPCITGSDI	120

Db 61 DPWPBLSIAKHEDITITAPGKERADSVLSALKALEDIASENDWVYHDAARPCLTGSDI 120
Qy 121 HLQIDTLKNDPVGGIILASSHDTLKHVGDGDTITATIDRKHWRAALTPOMKYKGLRDALQ 180
Db 121 HLQIDTLKNDPVGGIILASSHDTLKHVGDGDTITATIDRKHWRAALTPOMKYKGLRDALQ 180
Qy 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231
Db 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231

RESULT 2

US-09-934-868-68
Sequence 68, Application US/09934868
Patent No. US20020137190A1
GENERAL INFORMATION:
APPLICANT: Koffas, Mattheos
APPLICANT: Odum, James M.
APPLICANT: Schenzle, Andreas J.
TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN
FILE REFERENCE: CL1596 US NA
CURRENT APPLICATION NUMBER: US/09/934,868
CURRENT FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/229,858
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 81
SOFTWARE: Microsoft Office 97
SEQ ID NO 68
LENGTH: 231
TYPE: PRT
ORGANISM: Methylomonas 16a
FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by ISPD
US-09-934-868-68

Query Match 100.0%; Score 1199; DB 10; Length 231;
Best Local Similarity 100.0%; Pred. No. 2,1e-115;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNPTIQMAVVPAAVGKQADRPKQYLPAGKTYIEHTLTLLSSDAFQKVAVAISVE 60
Db 1 MNPTIQMAVVPAAVGKQADRPKQYLPAGKTYIEHTLTLLSSDAFQKVAVAISVE 60
Qy 61 DPWPBLSIAKHEDITITAPGKERADSVLSALKALEDIASENDWVYHDAARPCLTGSDI 120
Db 61 DPWPBLSIAKHEDITITAPGKERADSVLSALKALEDIASENDWVYHDAARPCLTGSDI 120
Qy 121 HLQIDTLKNDPVGGIILASSHDTLKHVGDGDTITATIDRKHWRAALTPOMKYKGLRDALQ 180
Db 121 HLQIDTLKNDPVGGIILASSHDTLKHVGDGDTITATIDRKHWRAALTPOMKYKGLRDALQ 180
Qy 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231
Db 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231

RESULT 3

US-09-941-947A-10
Sequence 10, Application US/09941947A
Publication No. US20030003528A1
GENERAL INFORMATION:
APPLICANT: Brzostowicz, Patricia C.
APPLICANT: Cheng, Qiong
APPLICANT: Dicosimo, Deana J.
APPLICANT: Koffas, Mattheos
APPLICANT: Miller, Edward S. Jr.
APPLICANT: Odum, J. Martin
APPLICANT: Picataggio, Steve
APPLICANT: Roviato, Pierre F.
TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
FILE REFERENCE: CL1903 US NA
CURRENT APPLICATION NUMBER: US/09/941,947A

CURRENT FILING DATE: 2001-09-01
FOR APPLICATION NUMBER: 60/229,907
FOR FILING DATE: 2000-09-01
FOR APPLICATION NUMBER: 60/229,858
FOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 60
SOFTWARE: Microsoft Office 97
SEQ ID NO 10
LENGTH: 231
TYPE: PRT
ORGANISM: Methylomonas 16a
US-09-941-947A-10

Query Match 100.0%; Score 1199; DB 11; Length 231;
Best Local Similarity 100.0%; Pred. No. 2,1e-115;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNPTIQMAVVPAAVGKQADRPKQYLPAGKTYIEHTLTLLSSDAFQKVAVAISVE 60
Db 1 MNPTIQMAVVPAAVGKQADRPKQYLPAGKTYIEHTLTLLSSDAFQKVAVAISVE 60
Qy 61 DPWPBLSIAKHEDITITAPGKERADSVLSALKALEDIASENDWVYHDAARPCLTGSDI 120
Db 61 DPWPBLSIAKHEDITITAPGKERADSVLSALKALEDIASENDWVYHDAARPCLTGSDI 120
Qy 121 HLQIDTLKNDPVGGIILASSHDTLKHVGDGDTITATIDRKHWRAALTPOMKYKGLRDALQ 180
Db 121 HLQIDTLKNDPVGGIILASSHDTLKHVGDGDTITATIDRKHWRAALTPOMKYKGLRDALQ 180
Qy 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231
Db 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231

RESULT 4

US-09-815-242-11931
Sequence 11931, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Olsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
THE OP INVENTION: Prokaryotes
FILE REFERENCE: BL17A.011a
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
FOR APPLICATION NUMBER: 60/191,078
FOR FILING DATE: 2000-03-21
FOR APPLICATION NUMBER: 60/206,848
FOR FILING DATE: 2000-05-23
FOR APPLICATION NUMBER: 60/207,727
FOR FILING DATE: 2000-05-26
FOR APPLICATION NUMBER: 60/242,578
FOR FILING DATE: 2000-10-23
FOR APPLICATION NUMBER: 60/253,625
FOR FILING DATE: 2000-11-27
FOR APPLICATION NUMBER: 60/257,931
FOR FILING DATE: 2000-12-32
FOR APPLICATION NUMBER: 60/269,308
FOR FILING DATE: 2001-02-16
FOR FILING DATE: 2001-14-10
NUMBER OF SEQ ID NOS: 1410
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 11931
LENGTH: 234
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11931

Query Match 46.3%; Score 555.5; DB 9; Length 234;
 Best Local Similarity 54.1%; Pred. No. 5.3e-49;
 Matches 119; Conservative 26; Mismatches 72; Indels 3; Gaps 3;

QY 8 AAVPAAGVGRMQRADPKQYPLAGKTVIEHTLTRLLBSDAFQKVAVAIVEDPYWPEL 67
 DB 10 WTVIPAGVGRMRADPKQYLDLAGRTVIERLDCFLHEPMLRGVLAVCLAEDDPYWPEL 69

QY 68 SIAGKPDITFAGKERADSVSA-LKALEDIASENDVAVHDAARPCLTGSDIHLQIDT 126
 DB 70 DCAARHVRQARAGKGRASVINGLRLELGAQDDVAVHDAARPCLTGSDIHLQIDT 129

QY 127 LKNDVGGILALSHDTLKHVDG-TTATIDRKHWRAALTPQWFKYGMRLDAL-QRTG 184
 DB 130 LAEDVGGILAVPADTLKESDRDGRVSEITDRSVVLAATPQWFKYGMRLDALALVA 189

QY 185 NPATDEASALELGHKPKIVEGRPDNITRPPDLALAQ 224
 DB 190 GVAITDEASALEMAGYAPRLVEGRADNITRPPDLALAQ 229

RESULT 5
 US-09-815-242-11076
 ; Sequence 11076, Application US/09815242
 ; Patent No. US2002061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: EITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11076
 ; LENGTH: 225
 ; TYPE: PRF
 ; ORGANISM: Haemophilus influenzae
 ; US-09-815-242-11076

Query Match 44.2%; Score 530.5; DB 9; Length 225;
 Best Local Similarity 52.5%; Pred. No. 1.9e-46;
 Matches 117; Conservative 28; Mismatches 65; Indels 13; Gaps 6;

QY 9 AAVPAAGVGRMQRADPKQYPLAGKTVIEHTLTRLLBSDAFQKVAVAIVEDPYWPEL 68
 DB 7 AAVPAAGVGRMQRADPKQYPLAGKTVIEHTLTRLLBSDAFQKVAVAIVEDPYWPEL 66

QY 69 IAKHPDITFAGKERADSVSA-LKALEDIASENDVAVHDAARPCLTGSDIHLQIDT 128
 DB 67 L--DEKIVGEGTTRAEVANGWA---IAEKNAWVTVHDAARPCLTGSDIHLQIDT 117

QY 129 --NDPVGGILALSHDTLKHVDG-TTATIDRKHWRAALTPQWFKYGMRLDALQRTG 184
 DB 118 AIBKQALIALPVTDLTKRADNQGCTVKTEDRSQLMQWTFPPFDILRDLSTGIDQ 177

QY 185 NPATDEASALELGHKPKIVEGRPDNITRPPDLALAQ 227
 DB 178 GANTDEASALELGHKPKIVEGRPDNITRPPDLALAQ 220

RESULT 6
 US-09-815-242-11722
 ; Sequence 11722, Application US/09815242
 ; Patent No. US2002061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: EITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11722
 ; LENGTH: 254
 ; TYPE: PRF
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-815-242-11722

Query Match 44.2%; Score 529.5; DB 9; Length 254;
 Best Local Similarity 51.1%; Pred. No. 2.9e-46;
 Matches 114; Conservative 30; Mismatches 70; Indels 9; Gaps 5;

QY 9 AAVPAAGVGRMQRADPKQYPLAGKTVIEHTLTRLLBSDAFQKVAVAIVEDPYWPEL 68
 DB 33 AAVPAAGVGRMQRADPKQYPLAGKTVIEHTLTRLLBSDAFQKVAVAIVEDPYWPEL 92

QY 69 IAKHPDITFAGKERADSVSA-LKALEDIASENDVAVHDAARPCLTGSDIHLQIDT 128
 DB 93 LAHPDITFAGKERADSVSA-LKALEDIASENDVAVHDAARPCLTGSDIHLQIDT 148

QY 129 NDPVGGILALSHDTLKHVDG-TTATIDRKHWRAALTPQWFKYGMRLDALQRTG 184
 DB 149 TSVGGILAVPADTLKESDRDGRVSEITDRSVVLAATPQWFKYGMRLDALALVA 208

QY 185 NPATDEASALELGHKPKIVEGRPDNITRPPDLALAQ 227
 DB 209 AATTDASALELGHKPKIVEGRPDNITRPPDLALAQ 250

RESULT 7
 US-09-815-242-10273

```

; Sequence 10273, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 10273
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-815-242-10273

Query Match      43.4%; Score 520.5; DB 9; Length 236;
Best Local Similarity 49.8%; Pred. No. 2.2e-45;
Matches 111; Conservative 31; Mismatches 72; Indels 9; Gaps 4;

QY 9 AVPAAGVGKRMQADRPKQYLPAGKTIVIEHTLTRLLESADAFQKVAVAISVEDPYMELS 68
DB 10 AVPAAGVGKRMQADRPKQYLPAGKTIVIEHTLTRLLESADAFQKVAVAISVEDPYMELS 69
QY 69 IAKHPDITTAAGKERADSVLSAKLEADIASENDVAVHDARPCLTGSDIHLQIDTLK 128
DB 70 LANHPQITVVDGDERADSVLAGLKA---AGDAQWLVHDARPCLTGSDIHLQIDTLK 125
QY 129 NDVPGIILALSHDTLKHVD--GDTTATIDRKHWRALTPQWKKYGMULDALQ--TEG 184
DB 126 TSRTGGIILAPVDTMKRAEPGNALAHYVDKNGLMHALTPQFFPRELHDCLTRALNEG 185
QY 185 NPATVDEASALELIGKRPKIVGSRPNIKITRPEDLALAQFYM 227
DB 186 -ATTDEASALEYCGFHPQLVGRADNIKVTREPDALAEFYL 227

RESULT 8
US-09-792-251-11
; Sequence 11, Application US/09792251
; Patent No. US20020160364A1
; GENERAL INFORMATION:
; APPLICANT: Fritzt, Christian
; APPLICANT: Youngman, Philip
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: USE OF YACM AND YOEJ, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 06286-140001
; CURRENT APPLICATION NUMBER: US/09/792,251
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 28

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 11
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-792-251-11

Query Match      43.4%; Score 520.5; DB 10; Length 236;
Best Local Similarity 49.8%; Pred. No. 2.2e-45;
Matches 111; Conservative 31; Mismatches 72; Indels 9; Gaps 4;

QY 9 AVPAAGVGKRMQADRPKQYLPAGKTIVIEHTLTRLLESADAFQKVAVAISVEDPYMELS 68
DB 10 AVPAAGVGKRMQADRPKQYLPAGKTIVIEHTLTRLLESADAFQKVAVAISVEDPYMELS 69
QY 69 IAKHPDITTAAGKERADSVLSAKLEADIASENDVAVHDARPCLTGSDIHLQIDTLK 128
DB 70 LANHPQITVVDGDERADSVLAGLKA---AGDAQWLVHDARPCLTGSDIHLQIDTLK 125
QY 129 NDVPGIILALSHDTLKHVD--GDTTATIDRKHWRALTPQWKKYGMULDALQ--TEG 184
DB 126 TSRTGGIILAPVDTMKRAEPGNALAHYVDKNGLMHALTPQFFPRELHDCLTRALNEG 185
QY 185 NPATVDEASALELIGKRPKIVGSRPNIKITRPEDLALAQFYM 227
DB 186 -ATTDEASALEYCGFHPQLVGRADNIKVTREPDALAEFYL 227

RESULT 9
US-09-815-242-13867
; Sequence 13867, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 13867
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE: VARIANT
; NAME/KEY: (1)....(236)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-815-242-13867

Query Match      41.7%; Score 499.5; DB 9; Length 236;

```


RESULT 13

US-10-156-761-11503
 ; Sequence 11503, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 11503
 ; LENGTH: 250
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-11503

Query Match 20.5%; Score 245.5; DB 15; Length 250;
 Best Local Similarity 32.1%; Pred. No. 5.7e-17;

Matches 75; Conservative 34; Mismatches 96; Indels 29; Gaps 6;

QY 9 AVPAAGVGMQADRPKQYLPAGKTVIEHTLTLLSDAFQKAVAI-----SV 59
 DB 16 AVIPAGRGVRIGCAPKALRALNGTPMLIHAVRAAPASLVVVAPEPDAEVS 75
 QY 60 EDPYWEISIAHPDITTPGKERADSVLSALKLEDSNDWVYHDAARPCLTGSD 119
 DB 76 LDH-----ALPSTDPVVVPGSSROESVKTGLDL-PPGIDVLYHDAARPLVPVDF 128
 QY 120 IHQIDTLKNDPVGGLIALSSHDTLKV-----DGDITITIDRKHWRALTPQMFYK 172
 DB 129 VDAIVEAVDGAFAVVPALPLADTVKQVAPAAVPGSEPVATPERARLRAVQTQ---- 184
 QY 173 GMLRDLQRTGKNPA--VTDEASALELGHKPKYVGRPDNITKTRPEDIALAQ 224
 DB 185 GFDRLTVRAHPTVDNVTDDASWEQIGARVVVPGHEBAPKTRPDLVLA 238

RESULT 14

US-10-289-762-621
 ; Sequence 621, Application US/10289762
 ; Publication No. US20040006218A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Griffiths, R.
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
 ; FILE REFERENCE: 9710-003-999
 ; CURRENT APPLICATION NUMBER: US/10/289,762
 ; CURRENT FILING DATE: 2003-03-27
 ; NUMBER OF SEQ ID NOS: 6849
 ; SEQ ID NO 621
 ; LENGTH: 218
 ; TYPE: PRT
 ; ORGANISM: Chlamydia pneumoniae
 US-10-289-762-621

Query Match 20.1%; Score 240.5; DB 12; Length 218;
 Best Local Similarity 32.9%; Pred. No. 1.5e-16;

Matches 74; Conservative 41; Mismatches 91; Indels 19; Gaps 9;

QY 3 PTICMAVVPAAVGMQADRPKQYLPAGKTVIEHTLTLLSDAFQKAVAISSVDP 62

DB 7 FMKSSLLISGGGGTRFGSKIPKQYLPINGTPLVHSL--KLSS--LPQIAEVIWCDP 63

QY 63 YWPEISIAKHEDDITTPGKERADSVLSALKLEDSNDWVYHDAARPCLTGSDIHL 122

DB 64 SYOR--TPOEYVSPAIR--GERRODSVESGLQOV-----SYPMVLIHGARPPIYDEIR- 115

QY 123 QIDTLKNDPVGGLIALSS--HDTLKHVDGDTTITIDRKHWRALTPQMFYKGLRALAQ 180

DB 116 --DLIETKXIGARLALSPITYTIKQNP--VNTLDRDNALAIHTPOCTITELREGLA 170

QY 181 -RTGNPAVDEASALELGHKPKYVGRPDNITKTRPEDIALAQ 224

DB 171 LAKGKQLTVNDIRAAHIIIGKPSQLVFNKHPQIKISYEDLTIQA 215

RESULT 15

US-09-738-626-6422
 ; Sequence 6422, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAOKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASANO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/158162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0
 ; SEQ ID NO 6422
 ; LENGTH: 256
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6422

Query Match 18.6%; Score 223.5; DB 10; Length 256;
 Best Local Similarity 30.7%; Pred. No. 1.1e-14;

Matches 71; Conservative 43; Mismatches 102; Indels 15; Gaps 7;

QY 9 AVPAAGVGMQADRPKQYLPAGKTVIEHTLTLLSDAFQKAVAISSVDPYWEPEL 67

DB 10 ALTLAAGRGTRIGPIPKAVTLRERTLERSIQAMULTSVDHIIIVSDMETYARDL 69

QY 68 ----SIKHDP---ITAPGKERADSVLSALKLE--DISENDWVYHDAARPCLTGSD 119

DB 70 LKRGGLNDPVGAVRLVHGGGEPDVSVMGLQISLDDPDPAIVLHDSARALTPPGM 129

QY 120 IHQIDTLKNDPVGGLIALSSHDTLKV-----DGDITITIDRKHWRALTPQMFYKGLRD 177

DB 130 IARVRYKHGATGAVIVLPVPSDTIKRVSDDGVVVTTPRARLRAVQTQGFLLSELVA 189

QY 178 ALQR--TEGNPAV--TDEASALELGHKPKYVGRPDNITKTRPEDIALAQ 224

DB 190 ANEKFPADPNPGITFDASLMEWYGADVVCVQGDPAFAFVTPIDIMLAAQ 240

Search completed: January 29, 2004, 16:21:16
 Job time : 16.7685 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 6.79678 Seconds
(without alignments)
3268.453 Million cell updates/sec

Title: US-09-941-947a-10

Perfect score: 1199

Sequence: 1 KMPTIGCAVAVPACVQKX.....IKTRPEDLALAQFVMEQQA 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	574.5	47.9	232	2	C82311
2	555.5	46.3	234	2	F83191
3	530.5	44.2	225	2	G64156
4	524.5	43.7	236	2	A91079
5	524.5	43.7	236	2	B85924
6	520.5	43.4	236	2	G65055
7	515.5	43.0	236	2	AE0856
8	512.5	42.7	241	2	AC0408
9	474.5	39.6	237	2	F84978
10	411.5	34.3	229	2	D81867
11	409.5	34.2	229	2	H81073
12	378	31.5	231	2	H82700
13	353.5	29.5	231	2	C83663
14	287	24.8	232	2	S66119
15	286	23.9	213	2	D70414
16	282.5	23.6	228	2	AG2451
17	279.5	23.3	400	2	A97535
18	279.5	23.3	400	2	AC2754
19	272	22.7	474	2	S60902
20	270	22.5	474	2	AD1104
21	268.5	22.4	232	2	B72259
22	263	21.9	230	2	AD1166
23	260.5	21.7	232	2	D70607
24	254.5	21.2	241	2	A86949
25	246	20.5	236	2	AP1210
26	241.5	19.6	390	2	A13359
27	234.5	19.1	211	2	H72061
28	234.5	19.6	211	2	H72061
29	234.5	19.6	211	2	H86562

30	233.5	19.5	229	2	P97291	4-diphosphocytidyl
31	227	18.9	237	2	AP1566	2-C-methyl-D-eryth
32	227	18.9	379	2	G34960	hypothetical prote
33	225	18.8	382	2	P87464	conserved hypotnet
34	214.5	17.9	238	2	B89789	conserved hypotnet
35	214.5	17.9	232	2	D75254	conserved hypotnet
36	209	17.4	235	2	D98015	conserved hypotnet
37	209	17.4	235	2	P95147	hypothetical prote
38	201.5	16.8	218	2	C61659	conserved hypotnet
39	200.5	16.7	238	2	P89768	hypothetical prote
40	190.5	15.9	399	2	G71314	conserved hypotnet
41	170.5	14.2	278	2	T00613	2-C-methyl-D-eryth
42	146.5	12.2	189	2	B71511	2-C-methyl-D-eryth
43	141	11.8	229	2	B71148	hypothetical prote
44	131.5	11.0	371	2	B81256	hypothetical prote
45	118	9.8	458	2	C68860	hypothetical prote

ALIGNMENTS

RESULT 1

C82311 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60) - Vibrio cholerae

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Mar-2003

C:Accession: C82311

R:Heidelberg, J.F.; Bisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Base, S.; Qin, H.; Dragoti, I.; Sellers, P.

1. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: C82311

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-232 <HEI>

A:Cross-references: GB:AE004139; GB:AE003852; NID:G9654953; PID:AA93696.1; GSPDB:GN001;

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

A:Genes: VC0528

A:Map position: 1

C:Suprafamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase

C:Keywords: nucleotidyltransferase

Query Match 47.9%; Score 574.5; DB 2; Length 232;
Best Local Similarity 54.4%; Pred. No. 4.5e-41;
Matches 124; Conservative 28; Mismatches 63; Indels 13; Gaps 5;

QY	9	AVVPAAGVGRVQADRPQYVPLAGKTYIETLRLESDFQKAVASVSDPVPPLS	68
DB	6	AIVPAAGVGRVQADRPQYVPLDQTYLERTVHLSEHIVAVASADDPFANLP	65
QY	69	IAKHPIITAPGKERADSVLSALKALEDIAS--NDVTLVHDARPLGSDHRLQIDT	126
DB	66	LAHHRVIRNDGGERADSVLS---ALEYVQHLSSEVTVHDARPCVTADITQLIT	122
QY	127	LENDPVGILALSSHDTKAYDG-DTITATIDRKHWRFALTPQFKYKMD---AQR	181
DB	123	ALAHPIGAILASPVDTKRGDHLQOIVHTVDRFALMALTPQFRASLEERLPAALQO	182
QY	182	TEGNPAVTVDEASALELHGKPKIVEGRPDNITKTRPEDLALAQFVMEQ	229
DB	183	---QVTTIDASAFEMRGKERALVAGRADNITKTRPEDLALAEFTLSR	227

RESULT 2

P8319- conserved hypothetical protein PA3633 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 17-Mar-2003

C:Accession: P8319

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri

adman, S.; Yan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, J.; Loty, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
 A/Reference number: A82950; MUID:20437337; PMID:10984043
 A/Accession: F83191
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-236 <STO>
 A/Cross-references: GB:AE004783; GB:AE004091; NID:G9949786; PIDN:AAG07021.1; GSPDB:GN001
 A/Experimental source: strain PA01
 C/Genetics:
 A/Genes: PA3633
 C/Superfamily: 4-diphosphocytidylyl-2-methyl-D-erythritol synthase

Query Match 46.3%; Score 555.5; DB 2; Length 234;
 Best Local Similarity 54.1%; Pred. No. 1.8e-39;
 Matches 119; Conservative 26; Mismatches 72; Indels 3; Gaps 3;

QY 8 AVVPAAGVGRKQADRPKQYLPPLAGKTVIEHTLRLLESDFQKAVVAISVEDPYPELS 67
 DB 10 WYVPAAGVGRKQADRPKQYLPPLAGKTVIEHTLRLLESDFQKAVVAISVEDPYPELS 69
 QY 68 SIAGHPDITAPGKERADSVLSA-LKALSDIASNDVTVLHDAARPCLTGSDIHLDITL 126
 DB 70 DCAASHVQRAAGARASVNLGLRLLEAGQADVTVLHDAARPNLTRODRLLEE 129
 QY 127 LKDDPGVGLIALSHDTLKHVDG-TITATIDRKHYRRLTQMFYGYMLDAL-QRTBG 184
 DB 130 LAEDPGVGLIALVAPARTLKRSDRGVSBTIDRSVWMLVYTPQMFGLALHRLADALVA 189
 QY 185 NPATDEASALELLGKPKIVEGRPNIKITRPELDALAQ 224
 DB 190 GVAITDEASALELLGKPKIVEGRPNIKITRPELDALAQ 229

RESULT 3
 G64156
 Hypothetical protein H10672 - Haemophilus influenzae (strain Rd KW20)
 C/Species: Haemophilus influenzae
 C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 17-Mar-2003
 C/Accession: G64156

R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Stitzley, R.; Liu, L.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995
 A/Authors: Gnehm, C.L.; McDonald, L.A.; Smali, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A./Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.
 A/Reference number: A64000; MUID:95350630; PMID:7542800
 A/Accession: G64156
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-225 <TRIG>
 A/Cross-references: GB:U32750; GB:U42023; NID:G1573668; PIDN:AAC2332.1; PID:G1573673; T
 C/Superfamily: 4-diphosphocytidylyl-2-methyl-D-erythritol synthase

Query Match 44.2%; Score 530.5; DB 2; Length 225;
 Best Local Similarity 52.5%; Pred. No. 2.2e-37;
 Matches 117; Conservative 28; Mismatches 65; Indels 13; Gaps 6;

QY 9 AVVPAAGVGRKQADRPKQYLPPLAGKTVIEHTLRLLESDFQKAVVAISVEDPYPELS 68
 DB 7 AVVPAAGVGRKQADRPKQYLPPLAGKTVIEHTLRLLESDFQKAVVAISVEDPYPELS 66
 QY 69 IAGHPDITAPGKERADSVLSA-LKALSDIASNDVTVLHDAARPCLTGSDIHLDITL 128
 DB 67 L-DPRTQIVGEGTAAESVINGLNA-LAEKNAAVTVLHDAARPCLTGSDIHLDITL 117
 QY 129 --NDPVGGLIALSHDTLKHVDG-TITATIDRKHYRRLTQMFYGYMLDAL-QRTBG 184
 DB 118 AIEDKQALIALVPTDITKADQOCIVTEDEKRSQMLQMTQPFVPTLRDLALSTG100 177

QY 185 NPATDEASALELLGKPKIVEGRPNIKITRPELDALAQYTM 227
 DB 178 GANITDEASALELLGKPKIVEGRPNIKITRPELDALAEYTL 220

RESULT 4
 A91079
 Hypothetical protein Ecs3601 [imported] - *Escherichia coli* (strain O157H7, substrain R10)
 C/Species: *Escherichia coli*
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 17-Mar-2003
 C/Accession: A91079
 R/Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasekura, N.; Yasunaga, T.; Kumara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA 88, 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic islands
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: A91079
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-236 <HAY>
 A/Cross-references: GB:BA000007; PIDN:BA97024.1; PID:G13363072; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain R10D 0509952
 C/Genetics:
 A/Genes: Ecs3601
 C/Superfamily: 4-diphosphocytidylyl-2-methyl-D-erythritol synthase

Query Match 43.7%; Score 524.5; DB 2; Length 236;
 Best Local Similarity 50.2%; Pred. No. 7.6e-37;
 Matches 112; Conservative 31; Mismatches 71; Indels 9; Gaps 4;

QY 9 AVVPAAGVGRKQADRPKQYLPPLAGKTVIEHTLRLLESDFQKAVVAISVEDPYPELS 68
 DB 10 AVVPAAGVGRKQADRPKQYLPPLAGKTVIEHTLRLLESDFQKAVVAISVEDPYPELS 69
 QY 69 IAGHPDITAPGKERADSVLSA-LKALSDIASNDVTVLHDAARPCLTGSDIHLDITL 128
 DB 70 LANPQITVVDKSGERADSVLSA-LKALSDIASNDVTVLHDAARPCLTGSDIHLDITL 125
 QY 129 NDPVGGLIALSHDTLKHVDG-TITATIDRKHYRRLTQMFYGYMLDAL-QRTBG 184
 DB 126 TSPGGLIALSHDTLKHVDG-TITATIDRKHYRRLTQMFYGYMLDAL-QRTBG 185
 QY 185 NPATDEASALELLGKPKIVEGRPNIKITRPELDALAQYTM 227
 DB 186 AAITDEASALELLGKPKIVEGRPNIKITRPELDALAEYTL 227

RESULT 5

B85924
 Hypothetical protein ygbp [imported] - *Escherichia coli* (strain O157H7, substrain EDL933)
 C/Species: *Escherichia coli*
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-Mar-2003
 C/Accession: B85924
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 111er, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: B85924
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-236 <STO>
 A/Cross-references: GB:AE005174; NID:G12517201; PIDN:AAG57854.1; GSPDB:GN00145; UMGF:240
 A/Experimental source: strain O157:H7, substrain BDL933
 C/Genetics:
 A/Genes: ygbp
 C/Superfamily: 4-diphosphocytidylyl-2-methyl-D-erythritol synthase

Query Match 43.7%; Score 524.5; DB 2; Length 236;
 Best Local Similarity 50.2%; Pred. No. 7.6e-37;
 Matches 112; Conservative 31; Mismatches 71; Indels 9; Gaps 4;

QY 9 AVVPAAGVGRKQADRPKQYLPPLAGKTVIEHTLRLLESDFQKAVVAISVEDPYPELS 68

```

Db      10 AVPAAGGRRMOTCECPKOYLSTGNQTLTSHSVALLAHPRVKRVVLAISGDSRFAQLP 69
      69 IAKHPDITAPGKERADSVLSALKALEDIASENDVVLVHDAARPCLTGSDIHLQIDTLK 128
      70 LANHQITVVDGGERADSVLAKLKA----AGDAQVVLVHDAARPCLTGSDIHLQIDTLK 125
      129 NDPVGGILALSSHDTLKAVD--GDTITATIDRKHWRAALTPOMKRYGMLDALQR--TEG 184
      126 TSTRGTGILAAPVDDTKKAEFGKNAIAHTVDRMGIMHALLTQFPFRELHDLCTRALNEG 185
      185 NPVATDEASALELGHKPKIVEGRPDNFKITRPEDLALAQPFM 227
      186 -ATTDEASALEYCGFHPQVVEGRADNFKITRPEDLALAEFTL 227

RESULT 6
65055
4-diphosphocytidyl-2C-methyl-D-erythritol synthase (EC 2.7.7.-) (similarity) - Escherich
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2003
C:Accession: G65055
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
-A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G65055
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-236 <31AT>
A:Cross-references: GB:AE000358; GB:U00096; NID:92367156; PID:AACT5789.1; PID:gl789104;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase
C:Keywords: nucleotidyltransferase

Query Match      43.4%; Score 520.5; DB 2; Length 236;
Best Local Similarity 49.8%; Pred. No. 1.7e-36;
Matches 111; Conservative 31; Mismatches 72; Indels 9; Gaps 4;

      9 AVPAAGGCKMADRPKOYLPLAKTVIENTLRLLESADAFQKAVAVISVEDPWPBLS 68
      10 AVPAAGGRRMOTCECPKOYLSTGNQTLTSHSVALLAHPRVKRVVLAISGDSRFAQLP 69
      69 IAKHPDITAPGKERADSVLSALKALEDIASENDVVLVHDAARPCLTGSDIHLQIDTLK 128
      70 LANHQITVVDGGERADSVLAKLKA----AGDAQVVLVHDAARPCLTGSDIHLQIDTLK 125
      129 NDPVGGILALSSHDTLKAVD--GDTITATIDRKHWRAALTPOMKRYGMLDALQR--TEG 184
      126 TSTRGTGILAAPVDDTKKAEFGKNAIAHTVDRMGIMHALLTQFPFRELHDLCTRALNEG 185
      185 NPVATDEASALELGHKPKIVEGRPDNFKITRPEDLALAQPFM 227
      186 -ATTDEASALEYCGFHPQVVEGRADNFKITRPEDLALAEFTL 227

RESULT 7
AB0856
2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (imported) - Salmonella enteric
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 17-Mar-2003
C:Accession: AB0856
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Mole, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Farry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AB0856
A:Status: preliminary

```

```

A:Molecule type: DNA
A:Residues: 1-236 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD06036.1; PID:gl6504003; GSPDB:GN00176
C:Genetics:
A:Gene: ybbp
C:Superfamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase

Query Match      43.0%; Score 515.5; DB 2; Length 236;
Best Local Similarity 49.8%; Pred. No. 4.4e-36;
Matches 111; Conservative 32; Mismatches 71; Indels 9; Gaps 4;

      9 AVPAAGGCKMADRPKOYLPLAKTVIENTLRLLESADAFQKAVAVISVEDPWPBLS 68
      10 AVPAAGGRRMOTCECPKOYLSTGNQTLTSHSVALLAHPRVKRVVLAISGDSRFAQLP 69
      69 IAKHPDITAPGKERADSVLSALKALEDIASENDVVLVHDAARPCLTGSDIHLQIDTLK 128
      70 LANHQITVVDGGERADSVLAKLKA----AGDAQVVLVHDAARPCLTGSDIHLQIDTLK 125
      129 NDPVGGILALSSHDTLKAVD--GDTITATIDRKHWRAALTPOMKRYGMLDALQR--TEG 184
      126 TSTRGTGILAAPVDDTKKAEFGKNAIAHTVDRMGIMHALLTQFPFRELHDLCTRALNEG 185
      185 NPVATDEASALELGHKPKIVEGRPDNFKITRPEDLALAQPFM 227
      186 -ATTDEASALEYCGFHPQVVEGRADNFKITRPEDLALAEFTL 227

RESULT 8
AC0406
4-diphosphocytidyl-2C-methyl-D-erythritol synthase (EC 2.7.7.-) (imported) - Yersinia pes
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-Mar-2003
C:Accession: AC0406
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
demo-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0406
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-241 <KUR>
A:Cross-references: GB:AL590842; PIDN:CA02591.1; PID:gl5981288; GSPDB:GN00175
C:Genetics:
A:Gene: ispD
C:Superfamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase
C:Keywords: nucleotidyltransferase

Query Match      42.7%; Score 512.5; DB 2; Length 241;
Best Local Similarity 49.1%; Pred. No. 8e-36;
Matches 111; Conservative 35; Mismatches 71; Indels 9; Gaps 4;

      9 AVPAAGGCKMADRPKOYLPLAKTVIENTLRLLESADAFQKAVAVISVEDPWPBLS 68
      13 AVPAAGIGRRMVDCEPKOYLTVGKTIHAFSLHPRIRVIVHPODQSRSL 72
      69 IAKHPDITAPGKERADSVLSALKALEDIASENDVVLVHDAARPCLTGSDIHLQIDTLK 128
      73 VADQPRISTYGGDQGRANSMAGIQ----LAQAEMVLVHDAARPCLTGSDIHLQIDTLK 128
      129 NDPVGGILALSSHDTLKAVD--GDTITATIDRKHWRAALTPOMKRYGMLDALQR--TEG 184
      129 CSQGGGILAPVDDTKKAEFGKNAIAHTVDRMGIMHALLTQFPFRELHDLCTRALNEG 188
      185 NPVATDEASALELGHKPKIVEGRPDNFKITRPEDLALAQPFM 230
      189 -VATDEASALEHCGHPIVLTGSRDNFKITRPEDLALAEFTLQR 233

RESULT 9
F84978

```

hypothetical protein [Imported] - Buchnera sp. (strain APS)
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-Mar-2003
C:Accession: F84978
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
N:ature 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A:Reference number: AB4930; MUID:20445173; PMID:1093077
A:Accession: F84978
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-237 <STO>
A:Cross-references: GB:AP000398; GSPDB:GNO0144
A:Experimental source: strain APS
C:Genetics:
A:Gene: yGBP; BU220
C:Superfamily: 4-diphosphocytidylyl-2-methyl-D-erythritol synthase

	Best Match	39.6%	Score 474.5;	DB 2	Length 237;
	Query Local Similarity	44.4%	Pred No. 1.3e-32;		
	Matches 99;	Conservative 50;	Mismatches 67;	Indels 7;	Gaps 3
Cy	9	A A V P A A G C G K Q A D P R K O Y L T P A G K T V I E H T L T L T L E S D A F O K A V A I S V E D P W P E L S	68		
Dd	13	A I V P A A G I G R M K I D V P K O Y I K I Q R T I L E H T L T L L P H N I V Q I I S L N K K O N Y P H K L S	72		
Cy	69	I A H P D I T T A P G K E A D S V T S A L K A B E D I A S E N D W T L Y H A A P C C L T G S D I H L Q I D T L K	128		
Dd	73	I S S E F R I I S V A G G E K R I N S L G L I V Y K V - - - D W I V E D A V P C S Y O L E K U L S I T I K	128		
Cy	129	N D P V G G I L A L S H D L T K H - - V D G D T I T A T I D E K H W A R A L P O M R K Y G M L D A L Q R - T E G N	185		
Dd	129	K N P G A L I A P A V S D I T K Y S N L K O K A V A V T Y V K R N L M H A L T Q L Q V E L L K N C L A K I K D Q	188		
Cy	186	P A Y T D E A S A L E L L G H R K P I V B E G R P N I K I T T R P E D A L A Q P T N E	228		
Dd	189	I S V T D E A S A L E Y G C V N P L L V S C R N K I T W E D I V L N P E F L K	231		

RESULT 10
D81867
Conserved hypothetical protein NMA1713 [Imported] - Neisseria meningitidis (strain Z2491)
C.Species: Neisseria meningitidis
C.Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 17-Mar-2003
C.Accession: D81867
R.Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morello,
R.; Holtroyd, S.; Jagals, K.; Leather, S.; Moule, S.; Mangall, K.; Quail, M.A.; Rajandream,
Mature 404, 502-506, 2000
A.Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A.Reference number: A81775; MUID:20222556; PMID:10761919
A.Accession: D81867
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-229 <PAR>
A.Cross-references: GB:A1162756; GB:A1157959; NID:G7380092; PIDN:CAB84941.1; PID:G738033
A.Experimental source: serogroup A, strain Z2491
C.Geneticos:
A.Gene: NMA1713
C.Superfamily: 4-diphosphocytidylyl-2-methyl-D-erythritol synthase

	Query Match	34.3%	Score 411.5;	DB 2;	Length 229;
	Best Local Similarity	39.7%	Pred. No. 2.5e-27;		
	Matches	91;	Conservative 48;	Mismatches 73;	Indels 17; Gaps 5
QY	9	A V P A A G V G K M O A R D P R K Y L T P L A G K T V E H T I L T L S D A F Q K A A V A S V E D P Y -----	63		
		: : :			
DG	7	A L I P A A G I G R R P A A D P P K O Y E T I G S T V D E H T I G I F R E H A V D L T V W V S P E D I P A D K Y G	66		
		: : :			
QY	64	- - W P E S I A X H P D I I T A P G K E R A D S V I S A L - K A L E - D I A S E N D M V I V H D A A P C I L T G S D	119		
		: : :			
DG	67	T A P Q Y R V W M N ----- G G T R A E T V R N G V A K L E T S I A L A E T D N I I L V H D A A R C U L P S E A	119		
		: : :			
QY	120	I H Q I D T L K A N D P V G I L A L S H D T L K H V D G D T I T I D R K V M T A L I P Q M E K Y G M L E A L	179		

Db 120 LRTLLIIGQAAAGGIGLIPAVDTIKCAGGNI SATYRTSTLMAQITPOLFRASLLIRAL 179

Qy 180 QRTGEPAYIDEASALZELGHPKRIVEGRPNIKITRDEDLAAQFYME 228

Db 180 -AAENIDGITDESAVERKGRIPLLVQGDANNTKLTQGDQDAIVIRLLLD 227

RESULT 11
H81073 conserved hypothetical protein NMB1513 [imported] - *Neisseria meningitidis* (strain MC58
C/Species: *Neisseria meningitidis*
C/Date: 31-Mar-2000 #sequence #revision 31-Mar-2000 #ext_change 17-Mar-2003
C/Accession: H81073
R/Retelink: H.; Saunders, N.J.; Heidelberg, J.; Joffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, C.; Fleischmann, R.D.; Dougherty, B.A.;
Li, H.; Qin, H.; Vamathavan, V.; Gill, J.; Scalapco, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vec-
A/Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A/Reference number: AB1000; MUID:20175755; PMID:10710307
A/Accession: H81073
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-229 <EXT>
A/Cross-references: GB:AE002501; GB:AE002098; NID:g7226755; PIDN:AAF41869.1; PID:g7226755
A/Experimental source: serogroup B, strain MC58
A/Genetics:
A:Gene: NMB1513
C:Superfamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase

Query Match	34.28	Score	409.5	DB 2	Length	229
Best Local Similarity	39.38	Pred.	3.7e-27			
Matches	90	Conservative	49	Mismatches	73	Indels
					17	Gaps
					5	

QY	9	A	V	P	A	G	A	G	K	K	O	A	D	R	P	K	O	T	L	P	L	A	G	K	T	V	I	E	T	L	L	R	L	E	S	D	A	F	O	K	N	A	V	A	S	V	E	D	P	Y	-----	63						
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
DB	7	A	L	P	A	A	I	G	R	F	G	A	D	K	R	O	Y	E	I	G	S	K	T	V	I	E	H	T	I	G	I	F	E	R	H	A	V	O	L	T	V	V	V	S	P	E	I	T	F	A	D	K	V	O	66			
QY	64	--	W	P	E	L	S	I	A	H	P	D	I	T	A	F	G	K	E	R	A	S	V	I	S	A	L	--	Y	A	L	E	--	D	I	S	E	N	D	M	V	L	H	E	A	R	P	C	T	G	S	D	119					
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
DB	67	T	A	P	O	V	A	V	M	K	-----	G	G	T	R	A	E	T	V	R	K	S	V	A	L	L	E	T	G	L	A	E	T	N	I	I	V	H	N	A	N	C	C	P	S	E	A	119										
QY	120	I	H	O	I	D	L	K	O	D	P	V	G	G	I	L	A	S	S	H	D	L	H	V	O	D	E	T	T	A	T	I	D	R	K	H	V	R	A	L	T	P	O	K	F	K	T	G	M	L	D	A	L	179				
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
DB	120	L	T	R	L	E	O	A	G	N	A	A	B	G	I	L	A	I	P	I	A	D	L	K	A	D	G	N	I	S	A	T	A	E	R	S	T	L	W	A	O	A	C	T	P	O	L	E	R	A	G	L	L	H	R	A	L	179
QY	180	C	R	E	G	N	A	V	T	D	E	A	S	A	L	E	L	I	G	H	K	P	I	V	A	R	G	R	D	N	I	K	T	R	P	E	L	A	L	O	A	F	M	E	228													
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
DB	180	A	A	E	N	L	D	I	T	D	E	A	S	A	V	K	L	G	V	P	R	L	I	O	D	V	A	R	N	L	K	T	P	O	A	T	I	V	R	L	L	I	D	227														

RESULT 12
 H82700
 conserved hypothetical protein XFI293 [imported] - *Xylella fastidiosa* (strain 955c)
 C/Species: *Xylella fastidiosa*
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Mar-2003
 C/Accession: H82700
 R/Anonymous: The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequences
 Nature 406, 151-157, 2000
 A>Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A/Reference number: AB2515: MIMD:20365717: PMID:10910447
 A/Note: for a complete list of authors see reference number A59328 below
 A/Accession: H82700
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-231 <STM>
 A/Cross-references: GB:AE003962; GB:AE003849; NID:G9106270; PIDM:AAF84102.1; GSPDB:GN001
 R:Experimental source: strain 955c
 R:Stimpos, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; A
 Rionas, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canaro, L.E.A.; Carraro, D.M.; Carer, H
 as-Neto, E.; Docena, C.; El-Destry, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froberg, J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Kieseer, J.E.; Kurnane, E.E.; Laigret, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Margues, M.V.; Martins, B.A.; Authors: Martins, E.M.B.; Matsumura, A.I.; Menck, C.F.M.; Oliveira, R.C.; Palmitieri, D.A.; Rodrigues, V.; Rosa, A.U. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasak, A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., M.A.; da Silveira, M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328
A:Contents: annotation
A:Genetics:
A:Gene: XF1293
C:Superfamily: 4-diphosphocytidylyl-2-methyl-D-erythritol synthase
Query Match 31.5%; Score 378; DB 2; Length 231;
Best Local Similarity 38.9%; Pred. No. 1.7e-24;
Matches 89; Conservative 37; Mismatches 101; Indels 2; Gaps 2;
Qy 4 TIOCAVPAAGVGRKQADPRKQYLPAGKTVIEHTITLLBESDAQVAIVASVEDPY 63
Db 2 SVGVAVIPAAAGRGVRFCSVPVKQYLPVAGRPILVYITLEAALPAVCGMLVVAABDLA 61
Qy 64 WPEISIAHPDITAPGGRADSVLSALKLEDIASENDWLVHDAARPCITGSDIHQ 123
Db 62 WSWTEVAGKRLVLCSSGVSTPAASVLSGLLAPGVHADPVLVHDAARVVALSDIERL 121
Qy 124 IDLEKNDPVGGILASSHDTLKHVDG- TITATIDRKVAFALTPQMFKGMEDALQRT 182
Db 122 LEACANHPVGGILAVPADITKRAAGDSIDGTEPRERLMAFPQCFRRSQVLVRGQVA 181
Qy 183 EGNP-AVDEASALELGHKPKIVEGRPDNFKITREPOLALAQFYMQQ 230
Db 182 AADGEITDEAMMERQGLRLVLCASNSKITTTPDVLVRFELAR 230

RESULT 13
C83663
hypothetical protein BH0107 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 17-Mar-2003
C:Accession: C83663
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Saeki, R.; Masui, N.; Fuji, F.; Hira Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C83663
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-228 <STO>
A:Cross-references: GB:AB001507; GB:BA000004; MIM:q10172612; PIRN:BA03826.1; GSPD3:GN000 A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0107
A:Superfamily: 4-diphosphocytidylyl-2-methyl-D-erythritol synthase
Query Match 29.5%; Score 353.5; DB 2; Length 228;
Best Local Similarity 37.0%; Pred. No. 1.9e-22;
Matches 84; Conservative 44; Mismatches 86; Indels 13; Gaps 6;
Qy 10 VPAAGVGRKQADPRKQYLPAGKTVIEHTITLLBESD-AFOQVAIVASVEDPYFELS 68
Db 6 VIPAAGGGRKARAGNKKQFIELGSKPIIAHTLAVFEDQDWCTNVVIANBOEIEEMELA 65
Qy 69 -----LAKPDITTPGGRKADSVLSALKLEDIASENDWLVHDAARPCITGSDIHQ 124
Db 66 NRYGISYAKKIIV--GGRERGESVPAIGKAL-----SQDGLVLIDHGARPPTKEIHSLV 119
Qy 125 DFLKNDPVGGILASSHDTLKHVDGTTITATIDRKVAFALTPQMFKGMEDALQSTEG 184
Db 120 ETAAKTH-AAVLAVPVDDITIKRVGEAVLETPPEELMAVOTPAFLALIKQAHQRAEN 178
Qy 185 NPAAV-TDEASALELGHKPKIVEGRPDNFKITREPOLALAQFYMQQ 230

[illegible]

V.

Mature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; KMD:98198566; PMID:9537320

A:Accession: D70414

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-213 <AOF>

A:Cross-references: GB:AE000734; NID:G2983733; PIDN:AA007307.1; PID:G2983746; GB:AE00065

A:Experimental source: strain VFS

A:Genetics:

A:Gene: aq_1323

C:Superfamily: 4-diphosphocytidylyl-2-methyl-D-erythritol synthase

Query Match 23.9%; Score 286; DB 2; Length 213;

Best Local Similarity 35.4%; Pred. No. 8, 7e-17;

Matches 79; Conservative 40; Mismatches 82; Indels 22; Gaps 8;

```
QY 9 AVVPAAGVGNQADRPQYLPAGKTVIERTLTLLSDFQKVAVAISVEDPYPELS 68
DB 4 AIIAAGSGSRI--GFRKQFATLCGKPLFMHSLERVL--DIFEEVILVL-----PEDF 52
QY 69 IAK--HPDIITAPGKXERADSVLSALALMEDIASENDVYVDAARPCLTGSDIHLQID 125
DB 53 LDKVKVHPKVKVAVAGGPERQDSVFNAL-----LQATGDIVIHDSARPLAT--KQMLEVA 106
QY 126 TKNDPVGCGILALSSHDTLKHVDGDTITATIDRKHWRALTPQWFKYGMLEADALQRTEN 185
DB 107 QL-GDYHGKVVASPARDTLKEVEGKVIKTLNRSIIWHAQTQAPFRDIIIECHBRAXAE 165
QY 186 PAV-TDEASALBELGKPKIVGGRPDNITKTRPEDLALAQFTM 227
DB 166 GFVGTDADSLERYSVGVGSGSYMNKITYPEDLEMYKKIM 208
```

Search completed: January 29, 2004, 15:55:30

Job time : 6.79678 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 4.34994 Seconds
(without alignments)
2497.314 Million cell updates/sec

Title: US-09-941-947A-10

Sequence: 1 MNPRTQCAAVYPAGVGKRM.....IKITRPDIALAGFYMQQA 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	605	50.5	237	1	ISPD_VIBVU	Q8d60 vibrio vuln
2	594.5	49.6	234	1	ISPD_VIBPA	Q871q2 vibrio para
3	574.5	47.9	232	1	ISPD_VIBCH	Q9KJ2 vibrio chol
4	555.5	46.3	234	1	ISPD_PSEAE	P57707 pseudomonas
5	530.5	44.2	225	1	ISPD_HAIAI	O08029 haemophilus
6	524.5	42.7	235	1	ISPD_ECO57	O87Y4 escherichia
7	520.5	43.4	235	1	ISPD_ECO16	O87E5 escherichia
8	520.5	43.4	235	1	ISPD_ECO1	Q4683 escherichia
9	515.5	43.0	236	1	ISPD_SALT1	O84771 salmonella
10	513.5	42.8	236	1	ISPD_SALT1	O84771 salmonella
11	512.5	42.7	241	1	ISPD_YERPE	O82b6 yersinia pe
12	505	42.1	238	1	ISPD_PASNU	P57953 pasteurella
13	474.5	39.6	237	1	ISPD_BUCAT	P57495 buchnera ap
14	462.5	38.6	229	1	ISPD_WIGBR	O84223 wiglesworth
15	457.5	38.2	236	1	ISPD_BUCAP	O839d6 buchnera ap
16	425	35.4	265	1	ISPD_XANCP	O85921 xanthomonas
17	422	35.2	266	1	ISPD_XANAC	O85178 xanthomonas
18	411.5	34.3	229	1	ISPD_NEIMA	O91tm3 neisseria m
19	409.5	34.2	229	1	ISPD_NEIMA	O91tm3 neisseria m
20	398	33.2	253	1	ISPD_RALSO	O8xyw3 ralsostonia s
21	378	31.5	231	1	ISPD_YLTRA	O9d6f8 xyloella fas
22	353.5	28.5	228	1	ISPD_BACHD	O9d6f8 xyloella fas
23	301	28.1	232	1	ISPD_BACHD	O9d6f8 xyloella fas
24	286	23.9	213	1	ISPD_AQUAE	O67343 aquifex aeo
25	282.5	23.6	228	1	ISPD_ANASP	O8y1x9 anabaena ap
26	279.5	23.3	400	1	ISPD_AGR75	O8y1f4 a ispd/ispf
27	270	22.5	232	1	ISPD_LISMO	O8y1f4 a ispd/ispf
28	268.5	22.4	222	1	ISPD_THENA	O8y1f4 a ispd/ispf
29	263	21.9	231	1	ISPD_FUSNA	O8y1f4 a ispd/ispf
30	263	21.9	232	1	ISPD_LISIN	O8y1f4 a ispd/ispf
31	260.5	21.7	230	1	ISPD_SYNY3	P74323 synechocyst
32	256	21.4	225	1	ISPD_CLOPE	O8xh3 clostridium
33	254.5	21.2	231	1	ISPD_MYCTU	P86864 mycobacteri

ALIGNMENTS

34	254.5	21.2	241	1	ISPD_MYCTE	O9cwe6 mycobacteri
35	247	20.6	227	1	ISPD_THETN	Q8786 thermotace
36	247	20.6	407	1	ISPD_RHTLO	Q98m9 streptomyce
37	243.5	20.3	270	1	ISPD_STRCO	Q910q8 streptomyce
38	241.5	20.1	390	1	ISPD_BRUME	O8yhd8 b ispd/ispf
39	234.5	19.6	211	1	ISPD_CHLPM	O92745 chlamydia p
40	233.5	19.5	229	1	ISPD_CLOAB	O976c9 clostridium
41	227	18.9	379	1	ISPD_RHOCA	O88113 r ispd/ispf
42	225	18.8	382	1	ISPD_CAOCR	O8a75 c ispd/ispf
43	225	18.8	434	1	ISPD_RHIME	O92q50 r ispd/ispf
44	223.5	18.6	256	1	ISPD_CORGL	O8nmb8 corynebacte
45	211	17.6	232	1	ISPD_DIRRA	O9r30 deinococcus

RESULT 1
ID ISPD_VIBVU STANDARD; PRT; 237 AA.
AC Q8d60;

DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE 2-C-methyl-D-erythritol 4-phosphate cyclidyltransferase (EC 2.7.7.60)
DE (4-diphosphocyclidyl-2C-methyl-D-erythritol synthase) (MSP
DE cyclidyltransferase) (MCT).
GN ISPD OR VY11582.

OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OX Vibrionaceae; Vibrio.
NCBI_TaxID=672;
NM NCBI_TaxID=672;
RM [1]

RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;

RT Submitted [DEC-2002] to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocyclidyl-2C-
methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-
phosphate (by similarity).
CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.

CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sb-sib.ch).

CC EMBL; AB016802; AA010005.1; -
DR HAVAP; MF_00108; -; 1.
DR InterPro; IPR001228; ISPD_synthase.
DR Pfam; PF01128; ISPD; 1.
DR TIGRFAMs; TIGR00453; ISPD; 1.
DR PROSITE; PS01295; ISPD; 1.
KW Transferase; Nucleotidyltransferase; Isoprene biosynthesis;
KW Complete proteome.
SQ ESQDENCE 237 AA; 26005 MM; 35D0952D12847BCD CRC64;

Query Match 50.5%; Score 605; DB 1; Length 237;
Best Local Similarity 55.4%; Pred. No. 9.6e-46;
Matches 124; Conservative 31; Mismatches 67; Indels 2; Gaps 2;

QY 9 AAVPAAGVGRMADREKQVLPAGKIVTERTLRLLSAPKQVAVISVEQPYEELS 68
DB 10 AAVPAAGVGRMADREKQVLPAGKIVTERTLRLLSAPKQVAVISVEQPYEELS 69

QY 69 IAKHDIITAPGKGRADSVLSALKALBEDIASENDWVWHDARPCLTGSDTHQIDTLK 128
 DB 70 IAKHDIITAPGKGRADSVLSALKALBEDIASENDWVWHDARPCLTGSDTHQIDTLK 129
 QY 129 NDPVGGIILASHTLKHVGD--TITATIDRKHWRAALTPQMEKYGMLDAL--QRTGPNP 186
 DB 130 SHETGGIILATPVDITMKRANAOQKIDHTYDRNALHMAALTPQMEKRAVLTLDALSLAQGV 189
 QY 187 ATDEASALIELGKPKIVGSRPNIKITRPEDLALAQFYMEQ 230
 DB 190 ATDEASALEMRGELPALVOGSSNIKYQPEDLALAEFYLSRE 233

RESULT 2

ISPD_VIBPA STANDARD; PRT; 234 AA.
 AC Q87I02;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (BC 2.7.7.60)
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP
 DE cytidyltransferase) (MCT).
 GN ISPD OR VP2559.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 OX NCBI_TaxID=6670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Naito M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shingawa H., Hatgori M., Iida T.,
 RA "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 RT distance from that of V. cholerae.";
 RL Lancet 361:743-749 (2003).
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-
 CC methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-
 CC phosphate (By similarity).
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AP005081; BAC60822.1; -
 DR HAMAP: MF_00108; -1
 DR PROSITE: PS01293; ISPD, 1.
 DR TRANSFERASE; Nucleotidyltransferase; Isoprene biosynthesis;
 DR Complete proteome.
 SQ SEQUENCE 234 AA; E30466C6D8C0682 CRC64;

Query Match 49.6%; Score 594.5; DB 1; Length 234;
 Best Local Similarity 55.9%; Pred. No. 7, 9e-45;
 Matches 127; Conservative 30; Mismatches 61; Indels 9; Gaps 4;

QY 9 AVVPAAGVGRKQADRPKQYLPAGKTVIEHTLTRLIESDAFQKVAVAISVEDPYPELS 68
 DB 10 AVVPAAGVGRKQADRPKQYLPAGKTVIEHTLTRLIESDAFQKVAVAISVEDPYPELS 69
 QY 69 IAKHDIITAPGKGRADSVLSALKALBEDIASENDWVWHDARPCLTGSDTHQIDTLK 128
 DB 70 IAKHDIITAPGKGRADSVLSALKALBEDIASENDWVWHDARPCLTGSDTHQIDTLK 128

QY 129 NDPVGGIILASHTLKHVGD--TITATIDRKHWRAALTPQMEKYGKMLDAL--QRTGPNP 183
 DB 129 AHPGSGIILASHTLKHVGD--TITATIDRKHWRAALTPQMEKYGKMLDAL--QRTGPNP 186
 QY 184 GNPVATDEASALIELGKPKIVGSRPNIKITRPEDLALAQFYMEQ 230
 DB 187 GNPVATDEASALEMRGELPALVOGSSNIKYQPEDLALAEFYLSRE 232

RESULT 3

ISPD_VIBCH STANDARD; PRT; 232 AA.
 AC Q9KUT2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (BC 2.7.7.60)
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP
 DE cytidyltransferase) (MCT).
 GN ISPD OR VC0528.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIIMD 2210633 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Rachev G., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Newman M.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.,
 RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.";
 RL Nature 406:477-483 (2000).
 CC -1- FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-
 CC METHYL-D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-
 CC PHOSPHATE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE004139; AAF93696.1; -
 DR PIR: C82311; C82311.
 DR HSSP: Q46893; 1152.
 DR TIGR: VC0528; -1
 DR HAMAP: MF_00108; -1
 DR InterPro: IPR001228; ISPD_synthase.
 DR Pfam: PF01128; ISPD, 1.
 DR PIRSF: PIRSF006765; DPCMB_synth; 1.
 DR TIGRPFAMs: TIGR00453; ISPD, 1.
 DR PROSITE: PS01293; ISPD, 1.
 DR TRANSFERASE; Nucleotidyltransferase; Isoprene biosynthesis;
 DR Complete proteome.
 SQ SEQUENCE 232 AA; 25901 MW; 44AA02420C724354 CRC64;

Query Match 47.8%; Score 574.5; DB 1; Length 232;
 Best Local Similarity 54.4%; Pred. No. 4, 4e-43;
 Matches 124; Conservative 28; Mismatches 63; Indels 13; Gaps 5;

QY 9 AVVPAAGVGRKQADRPKQYLPAGKTVIEHTLTRLIESDAFQKVAVAISVEDPYPELS 68

DB 6 AIVPAAGVGRKQADRPKQYLTLDKTYLLEHTVEHLLEHPLLEHVAVASADDFPFLALP 65
 QY 69 IAKEDPITTPGKGRKADSVLSALALEDIAS--NDVYLVDAAAPCLTGSDIHLQIDT 126
 DB 66 LAHFRVTRVNDGKERASVLS---ALBYVCOHRSEWLVVDAAAPCVTHADITQIIT 122
 QY 127 LKNDPVGGLIALSHDPLKRVDCD-TTATATIDRKVWRALTPQWFKYGMALD---ALOR 181
 DB 123 ALAHFVIGLIALSPVADPMKRGDHLQOITVITVDRKLMALHTPQWFRAGSLREHFLPALQ 182
 QY 182 TEGNPAVTDEASALELGHKPKIVEGRPNIKITRPEDIALAQFYMEQ 229
 DB 183 ---QVITDEASAFEMRGKFPALVAGRADNMLKITRPEDIALAEFLSR 227

RESULT 4

ISPD_PSEAE STANDARD; PRT; 234 AA.
 AC P5707; Q9HXZ7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP
 cytidyltransferase) (MCT).
 GN CYPIDY1LTRANSF (MCT).
 OS Pseudomonas aeruginosa.
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 NCBI_TaxID=287;
 RN [1]

SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAOI;
 PUBMED=10984043;
 MEDLINE=20437337;

RA Stover C.K., Phan X.-Q.T., Ertvin A.L., Mizoguchi S.D., Warren P.,
 Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
 Garber R.L., Goltzy L., Tolentino B., Westbrock-Wadman S., Yuan Y.,
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lam R.M.,
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 Reiter J., Sailer M.H., Hancock R.R.W., Lory Z., Olson M.V.,
 et al. Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an
 opportunistic pathogen.
 RA Nature 406:959-964 (2000).
 RL Nature 406:959-964 (2000).
 CC -1- FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-
 METHYL-D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-
 PHOSPHATE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 -1- diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; third step.
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; AB004783; AAC07021.1; -
 DR PIR; F83191; F83191.
 DR HSSP; Q46893; 1152.
 DR HAMAP; MF_00108; -; 1.
 DR InterPro; IPR001228; ISPD_synthase.
 DR Pfam; PF01128; ISPD; 1.
 DR PIRSF; PIRSF006765; DPCME_synth; 1.
 DR TIGRFAMs; TIGR00453; ISPD; 1.
 DR PROSITE; PS01295; ISPD; 1.
 KM Transferrase; Nucleotidyltransferase; Isoprene biosynthesis;
 KW Complete proteome.
 SQ SEQUENCE 234 AA; 25613 MF; 2CBA037CD85842C9 CRC64;

Query Match 46.3%; Score 555.5; DB 1; Length 234;
 Best local similarity 54.1%; Pred. No. 2e-41;
 Matches 119; Conservative 26; Mismatches 72; Indels 3; Gaps 3;

QY 8 WAVPAAGVGRKQADRPKQYLTPLAGTVEHTITRLLESDAFQKVAIVASVEDPYMBL 67
 DB 10 WTVPAAGVGRKQADRPKQYLTDLAGTVEHTITLDCLEHFMELGVLVCAIEDDPYMBL 69
 QY 68 SIAGHPDITTPGKGRKADSVLSA-LKALEDIASNDVYLVDAAAPCLTGSDIHLQIDT 126
 DB 70 DCAASHVQRAAGAEAGSVLSNGLRLLEAGQADVWLVDAAAPCLTGSDIHLQIDT 129
 QY 127 LKNDPVGGLIALSHDPLKRVDCD-TTATATIDRKVWRALTPQWFKYGMALD---QRTG 184
 DB 130 LAEDPVGGLAVPARDLTKSDRDGRVSEITIDRSVVLATTPQWFRGALRALADALVA 189
 QY 185 NPAVTDEASALELGHKPKIVEGRPNIKITRPEDIALAQ 224
 DB 190 GVAITDEASAFEMRGKFPALVAGRADNMLKITRPEDIALAEFLSR 229

RESULT 5

ISPD_HAEIN STANDARD; PRT; 225 AA.
 AC Q05029;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 23-FEB-2003 (Rel. 41, Last annotation update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP
 cytidyltransferase) (MCT).
 GN ISPD OR H10672.
 OS Haemophilus influenzae.
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 NCBI_TaxID=727;
 RN [1]

SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 Karlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 McCreney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 Fine L.D., Fritchman J.L., Fritchman J.L., Geoghagen N.S.M.,
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 Venter J.C., et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae*
 R1.
 RL Science 269:496-512 (1995).

CC -1- FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-
 D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-PHOSPHATE (BY
 SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 -1- diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; third step.
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; U32750; AAC22332.1; -
 DR PIR; G64156; G64156.
 DR HSSP; Q46893; 1152.
 DR TIGR; H10672; -
 DR HAMAP; MF_00108; -; 1.

DR InterPro: IPR001228; ISPD_synthase.
 DR Pfam: PF01128; ISPD; 1.
 DR PIRSF: PIRSF006765; DPCME_synth; 1.
 DR TIGRFAMs: TIGR00453; ISPD; 1.
 DR PROSITE: PS01295; ISPD; 1.
 DR Trasnase; Nucleotidyltransferase; Isoprene biosynthesis;
 KM Complete proteome.
 SQ SEQUENCE 225 AA; 24548 MW; 2155A297B44CD8C9 CRC64;
 Query Match 44.2%; Score 530.5; DB 1; Length 225;
 Best Local Similarity 52.5%; Pred. No. 2.9e-39;
 Matches 117; Conservative 28; Mismatches 65; Indels 13; Gaps 6;
 QY 9 AAVPAAGVGRMADRPKQYLPPLAGKTVVHTLRLLESADAPQKVAVAISVEDPYMELS 68
 DB 7 AAVPAAGVGRMADRPKQYLPPLAGKTVVHTLRLLESADAPQKVAVAISVEDPYMELS 66
 QY 69 IAKHPDITAPGGERADSVLSALKALBEDIASENDVAVDAARPCLTGSDIHLQIDTLK 128
 DB 67 L--DEPKIQLVEGGTTRAESEVLNGLNA---IAEKNAVAVHDAARPCLO---HADIDKLL 117
 QY 129 --NDPVGGLIASHDITLKVHVG-DITATIDRKHWREALTPQMKYGLSDALQOR-TEG 184
 DB 118 AIEDKQAGLALPVTDTIRGADNQCIVKTESGQLQWQMPFPVVDILADALSTGIQ 177
 QY 185 NPAVTDEASALBELGKPKIVBGRPNIKITRPEDLALAQFYM 227
 DB 176 GANITDEASALBELGKPKIVBGRPNIKITRPEDLALAQFYL 220
 RESULT 6
 ISPD_ECO57 STANDARD; PRT; 235 AA.
 AC 08x774;
 DT 28-SEP-2003 (Rel. 41; Last sequence update)
 DT 28-SEP-2003 (Rel. 41; Last sequence update)
 DT 15-SEP-2003 (Rel. 42; Last sequence update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyllyltransferase (EC 2.7.7.60)
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP
 DE cytidyllyltransferase) (MCT) (CDP-ME synthetase).
 GN ISPD OR Z4055 OR ECS3601.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller J.,
 RA Grobeck E.J., Davis N.W., Lim A., Ditalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H., Ogasawara N., Yasunaga T.,
 RA Kubera S., Shiba T., Hattori M., Shinagawa H., Ogasawara N., Yasunaga T.,
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:111-223 (2001).
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-methyl-
 D-erythritol from CDP + 2C-methyl-D-erythritol 4-phosphate =
 CC -1- CATALYTIC ACTIVITY: CDP + 2-C-methyl-D-erythritol 4-phosphate =
 CC -1- diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -1- COFACTOR: Magnesium, manganese or cobalt (By similarity).

CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EXBL; AED05502; AAG57854.1;
 DR EXBL; AP002562; BA837024.1;
 DR PIR; A91079; A91079.
 DR PIR; B85924; B85924.
 DR HAMAP; MP_00108; -- 1.
 DR InterPro: IPR001228; ISPD_synthase.
 DR Pfam; PF01128; ISPD; 1.
 DR PIRSF; PIRSF006765; DPCME_synth; 1.
 DR TIGRFAMs; TIGR00453; ISPD; 1.
 DR PROSITE; PS01295; ISPD; 1.
 DR Trasnase; Nucleotidyltransferase; Isoprene biosynthesis; Magnesium;
 KM Complete proteome.
 FT INIT MET 0
 SQ SEQUENCE 235 AA; 25602 MW; 3909048B4842ALDC CRC64;
 Query Match 43.7%; Score 524.5; DB 1; Length 235;
 Best Local Similarity 50.2%; Pred. No. 1e-38;
 Matches 112; Conservative 31; Mismatches 71; Indels 9; Gaps 4;
 QY 9 AAVPAAGVGRMADRPKQYLPPLAGKTVVHTLRLLESADAPQKVAVAISVEDPYMELS 68
 DB 9 AAVPAAGVGRMADRPKQYLPPLAGKTVVHTLRLLESADAPQKVAVAISVEDPYMELS 68
 QY 69 IAKHPDITAPGGERADSVLSALKALBEDIASENDVAVDAARPCLTGSDIHLQIDTLK 128
 DB 69 IAKHPDITAPGGERADSVLSALKALBEDIASENDVAVDAARPCLTGSDIHLQIDTLK 124
 DB 69 IAKHPDITAPGGERADSVLSALKALBEDIASENDVAVDAARPCLTGSDIHLQIDTLK 124
 QY 129 NDPVGGLIASHDITLKVHVG-DITATIDRKHWREALTPQMKYGLSDALQOR-TEG 184
 DB 125 TSRTGGLIAPVDTIRGADNQCIVKTESGQLQWQMPFPVVDILADALSTGIQ 184
 QY 185 NPAVTDEASALBELGKPKIVBGRPNIKITRPEDLALAQFYM 227
 DB 185 NPAVTDEASALBELGKPKIVBGRPNIKITRPEDLALAQFYL 226
 RESULT 7
 ISPD_ECO57 STANDARD; PRT; 235 AA.
 AC 08x774;
 DT 15-SEP-2003 (Rel. 42; Last sequence update)
 DT 15-SEP-2003 (Rel. 42; Last sequence update)
 DT 15-SEP-2003 (Rel. 42; Last sequence update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyllyltransferase (EC 2.7.7.60)
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP
 DE cytidyllyltransferase) (MCT).
 GN ISPD OR C3314.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22386234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Mayhew G.F., Rose D.J., Zhou S., Boutin A., Hackett J., Stroud D.,
 RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of enterohaemorrhagic Escherichia coli O157:H7";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).

CC - FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-
 CC methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-
 CC phosphate (By similarity).
 CC - CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 CC - diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC - COFACTOR: Magnesium, manganese or cobalt (By similarity).
 CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
 CC - SIMILARITY: BELONGS TO THE ISPD FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AB016765; AAN81763.1; -.
 DR HAMAP: MF_00108; -; 1.
 DR InterPro: IPR001228; ISPD_synthase.
 DR Pfam: PF01128; ISPD; 1.
 DR TIGRFAMs: TIGR00453; ISPD; 1.
 DR PROSITE: PS01295; ISPD; 1.
 DR Transferrase: Nucleotidyltransferase: Isoprene biosynthesis; Magnesium;
 KM Manganese, Cobalt; Complete proteome.
 FT INIT MET 0 By SIMILARITY.
 SQ SEQUENCE 235 AA; 25648 MW; A4DD31D950D6CE6F CRC64;
 Query Match 43.4%; Score 520.5; DB 1; Length 235;
 Best local Similarity 49.8%; Pred. No. 2,3e-38;
 Matches 111; Conservative 32; Mismatches 71; Indels 9; Gaps 4;
 QY 9 AAVPAAGCKRQADRPKQYLPAGKTVIEFTLRLLSDSAFOKAAVAISVDPYMPIS 68
 DB 9 AAVPAAGCKRQADRPKQYLPAGKTVIEFTLRLLSDSAFOKAAVAISVDPYMPIS 68
 QY 69 IAKHPDIIAPGKKEADSVLSAKALBEDIASENMVLVHDARCLGSDIHLIDTLK 128
 DB 69 LANHPITVDDGSEADSVLSAGLRA---AGDAQVVLVHDARPCLODDARLLALSE 124
 QY 129 NDVPGIALLSHDILKAVD--GDITATIDRHWVRAITPQFKYGMRLAQOR--TEG 184
 DB 125 TSRTGILAAPVDITKRAEPGKMAIAHTVDNGLMHALTPQFFREILHDLCTALNEG 184
 QY 185 NPAYVDEASALELGGKIVGRPNITCTRPEDLALQAFM 227
 DB 185 -ATTDEASALEYCGHPQLVEGRADNIKVTREPDLALAEFL 226
 RESULT 8
 ID ISPD_ECOLI STANDARD; PRT; 235 AA.
 AC Q46893;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DS 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MBP
 GN cytidyltransferase) (MCT) (CDP-ME synthetase).
 OS ISPD OR B2747.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RC STRAIN=K12 / DHS-alpha;
 RX MEDLINE=99449755; PubMed=10518523;
 RA Rohdich F., Wungstintaweekul J., Fellemeier M., Sagner S., Herz S.,
 RA Kls K., Bisenreich M., Bacher A., Zenk M.H.;
 RT "Cytidine 5'-triphosphate-dependent biosynthesis of isoprenoids: YgBP
 RT protein of Escherichia coli catalyzes the formation of 4-
 RT diphosphocytidyl-2-C-methylerythritol.";

RL Proc. Natl. Acad. Sci. U.S.A. 96:11758-11763(1999).
 RM [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Kuzayama T., Takagi M., Kaneda K., Dairi T., Seto H.;
 RT "Formation of 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol from
 RT 2-C-methyl-D-erythritol 4-phosphate by 2-C-methyl-D-erythritol
 RT 4-phosphate cytidyltransferase, a new enzyme in the nonmevalonate
 RT pathway.";
 RL Tetrahedron Lett. 41:703-706(2000).
 RM [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatterer F.R., Plunkett G. III, Bloch C.A., Perma N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RM [4]
 RP X-RAY CRYSTALLOGRAPHY (1.50 ANGSTROMS).
 RX MEDLINE=21320925; PubMed=11427897;
 RA Richard S.B., Bowman M.E., Kwiatkowski W., Kang I., Chow C.,
 RA Hillo A.W., Cane D.B., Noel T.P.;
 RT "Structure of 4-diphosphocytidyl-2-C-methylerythritol synthetase
 RT involved in mevalonate-independent isoprenoid biosynthesis.";
 RL Nat. Struct. Biol. 8:641-648(2001).
 CC - FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-
 CC D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-PHOSPHATE.
 CC - CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 CC - diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC - COFACTOR: MAGNESIUM, MANGANESE OR COBALT.
 CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
 CC - SUBUNIT: Homodimer.
 CC - MISCELLANEOUS: OPTIMAL PH IS 8.3.
 CC - SIMILARITY: BELONGS TO THE ISPD FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF230736; AAF43207.1; -.
 DR EMBL: AB037143; BAA90761.1; -.
 DR EMBL: U29579; AAG69257.1; -.
 DR EMBL: AB000358; AAC75789.1; -.
 DR PIR: G65055; G65055.
 DR PDB: 1I52; 1I-JUL-01.
 DR PDB: 1INI; 14-NOV-01.
 DR PDB: 1INU; 11-JUL-01.
 DR EcoGene: EG1110; ISPD.
 DR HAMAP: MF_00108; -; 1.
 DR InterPro: IPR001228; ISPD_synthase.
 DR Pfam: PF01128; ISPD; 1.
 DR PIRSF: PIRSF006765; DPCOE_synch; 1.
 DR TIGRFAMs: TIGR00453; ISPD; 1.
 DR PROSITE: PS01295; ISPD; 1.
 DR Transferrase: Nucleotidyltransferase: Isoprene biosynthesis; Magnesium;
 KM Manganese, Cobalt; 3D-structure; Complete proteome.
 FT INIT MET 0
 FT STRAND 7 13
 FT HELIX 18 20
 FT HELIX 26 28
 FT STRAND 30 31
 FT TURN 32 33
 FT STRAND 34 35
 FT HELIX 36 45
 FT TURN 46 46
 FT TURN 48 49

FT STRAND 50 57
 FT TURN 59 60
 FT HELIX 59 60
 FT HELIX 66 66
 FT HELIX 68 71
 FT TURN 73 74
 FT STRAND 75 79
 FT HELIX 84 93
 FT TURN 94 94
 FT TURN 96 97
 FT STRAND 100 103
 FT TURN 106 107
 FT TURN 109 110
 FT HELIX 113 120
 FT HELIX 121 124
 FT TURN 125 125
 FT STRAND 130 135
 FT TURN 140 143
 FT TURN 145 146
 FT STRAND 150 154
 FT TURN 157 158
 FT STRAND 160 169
 FT HELIX 170 182
 FT TURN 183 184
 FT HELIX 190 196
 FT TURN 197 198
 FT STRAND 202 205
 FT TURN 208 209
 FT HELIX 216 226
 SQ SEQUENCE 235 AA; 25606 MW; 2C090B31E2E81E6C CRC64;

Query Match 43.4%; Score 520.5; DB 1; Length 235;
 Best Local Similarity 49.8%; Pred. No. 2,3e-38;
 Matches 111; Conservative 31; Mismatches 72; Indels 9; Gaps 4;

QY 9 AVPAAGVGRMQRQADREKQYPLAGKTVIEHTLRLLESADAFQKVAVALSVEDPYWPELS 68
 DB 9 AVPAAGVGRMQRQADREKQYPLAGKTVIEHTLRLLESADAFQKVAVALSVEDPYWPELS 68
 QY 69 IAHNPITLAPGKERADSVLSALKLEDLASNDVYLVDAAAPCLTGSIHQIDTLK 128
 DB 69 IAHNPITLAPGKERADSVLSALKLEDLASNDVYLVDAAAPCLTGSIHQIDTLK 128
 QY 129 NDPVGGIILASHTLKHVD--GDITATIDRKHWRAITPOMFKYGMRLDALQR--TEG 184
 DB 125 TSTGGIILAPVDTMKRAEPGKNAIAHTVDNGMLHALLTPQFFRELLHDLCLTALNMG 184
 QY 185 NPAVTDEASALELGHKPKIVBGRPNVITKTRPEDIALAQTMM 227
 DB 185 -ATTIDEASALEYCGFHPALVGRADNIVKTRPEDIALAETYL 226

RESULT 9
 ID ISPD_SALTY STANDARD; PRT; 236 AA.

AC Q82471, 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC 2.7.7.60)
 DE (4-phosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP
 DE cytidylyltransferase) (MCT)
 GN ISPD OR STY3055 OR T2831.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_Taxid=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Cornerton P.,

RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Parrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Hoiyod S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.,
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RM [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Ty2 / ATCC 700931;
 RC MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.,
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -I- FUNCTION: Catalyzes the formation of 4-diphosphocytidylyl-2C-
 CC methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-
 CC phosphate (By similarity).
 CC -I- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -I- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
 CC -I- SIMILARITY: BELONGS TO THE ISPD FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; AL627276; CAD06036.1;
 DR EMBL; AE016843; AAC070388.1;
 DR HAMAP; MF 00108; -; 1.
 DR InterPro; IPR001228; ISPD_synthase.
 DR Pfam; PF01128; ISPD; 1.
 DR PIRSF; PIRSF006765; DPCME_synth; 1.
 DR TIGRFAMs; TIGR00453; ISPD; 1.
 DR PROSITE; PS01295; ISPD; 1.
 DR KX Transferase; Nucleocytidylyltransferase; Isoprene biosynthesis;
 KM Complete proteome.
 SQ SEQUENCE 236 AA; 25729 MW; DE4C81315483A90E CRC64;

Query Match 43.0%; Score 515.5; DB 1; Length 236;
 Best Local Similarity 49.8%; Pred. No. 6.4e-38;
 Matches 111; Conservative 32; Mismatches 71; Indels 9; Gaps 4;

QY 9 AVPAAGVGRMQRQADREKQYPLAGKTVIEHTLRLLESADAFQKVAVALSVEDPYWPELS 68
 DB 10 AVPAAGVGRMQRQADREKQYPLAGKTVIEHTLRLLESADAFQKVAVALSVEDPYWPELS 69
 QY 69 IAHNPITLAPGKERADSVLSALKLEDLASNDVYLVDAAAPCLTGSIHQIDTLK 128
 DB 70 IAHNPITLAPGKERADSVLSALKLEDLASNDVYLVDAAAPCLTGSIHQIDTLK 125
 QY 129 NDPVGGIILASHTLKHVD--GDITATIDRKHWRAITPOMFKYGMRLDALQR--TEG 184
 DB 126 NSRVGGIILASVYRTMKRGEKNAIAHYERADNIALTPQFFRELLHDLCLTALNMG 185
 QY 185 NPAVTDEASALELGHKPKIVBGRPNVITKTRPEDIALAQTMM 227
 DB 186 -ATTIDEASALEYCGFHPALVGRADNIVKTRPEDIALAETYL 227

RESULT 10
 ID ISPD_SALTY STANDARD; PRT; 236 AA.

AC Q82MF6, 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP
 DE cytidyltransferase) (MCT).
 GN ISPD OR STM2930.
 OC Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OC NCBI_TaxID=602;
 OK RN
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Girelli N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RT Nature 413:852-856 (2001).
 RL
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidylyl-2C-
 CC methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-
 CC phosphate (by similarity).
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE008833; AAL21810.1; -.
 DR StyGene; SGR7777; ISPD.
 DR HAMAP: MF_00108; -; 1.
 DR InterPro: IPR001228; ISPD_synthase.
 DR Pfam: PF01128; ISPD; 1.
 DR PIRSF: PIRSF006765; DPCKE_synth; 1.
 DR TIGRFAMs: TIGR00453; ISPD; 1.
 DR PROSITE: PS01295; ISPD; 1.
 DR Transferrase; Nucleotidyltransferase; Isoprene biosynthesis;
 DR Complete proteome.
 KW
 KW COMPLETE PROTEOME.
 SQ SEQUENCE 236 AA; 25755 MW; C4E631315492B80B CRC64;
 Query Match 42.8%; Score 513.5; DB 1; Length 236;
 Best Local Similarity 49.8%; Pred. No. 9,6e-38;
 Matches 111; Conservative 32; Mismatches 71; Indels 9; Gaps 4;
 QY 9 AVPAAGVGKQADRPKQYLPAGKTVIEHTLRLBSAFOKVAVAISVEDPYWPELS 68
 DB 10 AVPAAGFGRFQTECPKQYLSIGKTTIEHSVALLHAPRTAVVIAISPDHFAQLP 69
 QY 69 IAKHPDITTAPEGKERADSVLSALKALEDIASENDWVYVHDAARCLTSGDHLQIDTLK 128
 DB 70 IANHQITVVDGNERADSVLGLQAV-----AKAOWVHDAARCLTQDDIARLALISE 125
 QY 129 NDPVGGIILASSHDLKHYD--GDTITATIDRKHWARLALTPOMFKYGMILDALQR--TEG 184
 DB 126 NERVGGIILASPVADTMKREPERKNAIAHVERADIMHNLTPFPFRELIDLTITALNEG 185
 QY 185 NEAVTDEASALALLHKEKTVIEGRPDNKTITREPDLALAQPTM 227
 DB 186 -ATITDEASALIEYCGFHPALVEGRADNIIKVTREPDLALAEFYL 227
 RESULT 11
 ID ISPD_YERPE STANDARD; PRT; 241 AA.
 AC Q8ZBP6;

DT 23-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 23-FEB-2003 (Rel. 41, Last annotation update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP
 DE cytidyltransferase) (MCT).
 GN ISPD OR YPO3361 OR Y0828.
 OC Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OK NCBI_TaxID=632;
 OK RN
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parham J.J., Wren B.W., Thomson N.R., Tildall R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher G., Mungall K.L.,
 RA Baker S., Baaham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RA "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RT Nature 413:523-527 (2001).
 RL
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidylyl-2C-
 CC methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-
 CC phosphate (by similarity).
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AJ414156; CAC92591.1; -.
 DR EMBL: AE013685; AAM64413.1; -.
 DR PIR: AC0408; AC0408.
 DR HAMAP: MF_00108; -; 1.
 DR InterPro: IPR001228; ISPD_synthase.
 DR Pfam: PF01128; ISPD; 1.
 DR PIRSF: PIRSF006765; DPCKE_synth; 1.
 DR TIGRFAMs: TIGR00453; ISPD; 1.
 DR PROSITE: PS01295; ISPD; 1.
 DR Transferrase; Nucleotidyltransferase; Isoprene biosynthesis;
 DR Complete proteome.
 KW
 KW COMPLETE PROTEOME.
 SQ SEQUENCE 241 AA; 26401 MW; EF4CA6846BDC261 CRC64;
 Query Match 42.7%; Score 512.5; DB 1; Length 241;
 Best Local Similarity 49.1%; Pred. No. 1.2e-37;
 Matches 111; Conservative 35; Mismatches 71; Indels 9; Gaps 4;
 QY 9 AVPAAGVGKQADRPKQYLPAGKTVIEHTLRLBSAFOKVAVAISVEDPYWPELS 68
 DB 13 AVPAAGISRLAVDCRQYLVGKTTIEHAIISLHHRIRIGRYIVVHPODQFRLS 72
 QY 69 IAKHPDITTAPEGKERADSVLSALKALEDIASENDWVYVHDAARCLTSGDHLQIDTLK 128

Db 73 VADPRLSTVGGDDQGANSMAGLQ----LAGAENVIVHADARPCILHDDLSRLISTE 128
 QY 129 NDVVGGLIALSSDITLKHVDG--DTTATIDRKHWRPALTPOFKGKTGTDALQRT--EG 184
 Db 129 CSQVGGILAPVDVIMKRAEPGIAIAFTVDRQDLWALTPQLFPELTKLCLSRALREG 188
 QY 185 NPATVEASALBELGKPKIVEGSPDNKIKTRPEDIALAQFYEQ 230
 Db 189 -VAVTDEASALEHCGHPIILVTRSDNIKTRPEDIALAFYLTQR 233

RESULT 12
 ISPD_PASMT STANDARD; PRT; 238 AA.
 ID ISPD_PASMT
 AC P57953;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP
 cytidyltransferase) (MCT).
 GN ISPD OR PM1608
 OS Pasteurella multocida
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70.
 RX MEDLINE=2115866; PubMed=11248100;
 RA May 5.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida PM70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
 CC -1- FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-
 METHYL-D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-
 PHOSPHATE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -1- PATHWAY: Nonnevalonate terpenoid biosynthesis pathway; third step.
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL, AF006198; AK03692.1; -
 DR HSSP; Q46893; I152.
 DR HAMAP; MF_00108; -; 1.
 DR InterPro; IPR001228; ISPD_synthase.
 DR Pfam; PF01128; ISPD; 1.
 DR PIRSF; PIRSF006765; DPCME synth; 1.
 DR TIGRFAMs; TIGR00453; ISPD; 1.
 DR PROSITE; PS01295; ISPD; 1.
 DR Transferrase; Nucleotidyltransferase; Isoprene biosynthesis;
 KW Complete proteome.
 SQ SEQUENCE 238 AA; 26195 MW; 6AF21525CB4CFE22 CRC64;

Query Match 42.1%; Score 505; DB 1; Length 238;
 Best Local Similarity 47.6%; Pred. No. 5, 4e-37;
 Matches 107; Conservative 33; Mismatches 79; Indels 6; Gaps 4;

QY 9 AVPAAGVGKMQADRPQYIPLAGKTVIEHTLTRLESDFQKVAVAISVEDPYPELS 68
 Db 14 AVPAAGIGSRMQMDPKQYILHGTILIEHTLSVGLGPILEKILAAVADPYISTCP 73
 QY 69 IAKHPDITAPGGERADSVLSALKALEDA--SEWDVLYVHDAAPPCLTGSDIHQIDFLK 126
 Db 74 LITHPKIQVLEGGSSADSVLNGINAVKAVONSEDFWVWVHDAAPPCLTGDDKLVGV 133

QY 127 LKNPVGGLIALSSDITLKH--VDGTTATIDRKHWRPALTPOFKGKTGTDALQRTG 185
 Db 124 --EDNKGAILIAPDVTIKRALHNOQHITRDSQMLAQTPQFPPIATLQAALQALQA 191
 QY 186 P-AVDEASALELGHKPKIVEGSPDNKIKTRPEDIALAQFYEQ 229
 Db 192 GLQVTDASAMEFAFRPHLVAGSDNIKTRPEDIALAFYLSR 236

RESULT 13
 ISPD_BUCAL STANDARD; PRT; 237 AA.
 ID ISPD_BUCAL
 AC P57455;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP
 cytidyltransferase) (MCT).
 GN ISPD OR BU420.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 symbiotic bacterium)
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OX NCBI_TaxID=118099;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998.
 RX MEDLINE=2045173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 Buchnera sp. APS.";
 RL Nature 407:81-86 (2000).
 CC -1- FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-
 METHYL-D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-
 PHOSPHATE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -1- PATHWAY: Nonnevalonate terpenoid biosynthesis pathway; third step.
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL, AP001119; BAB13118.1; -
 DR HSSP; Q46893; I152.
 DR HAMAP; MF_00108; -; 1.
 DR InterPro; IPR001228; ISPD_synthase.
 DR Pfam; PF01128; ISPD; 1.
 DR PIRSF; PIRSF006765; DPCME synth; 1.
 DR TIGRFAMs; TIGR00453; ISPD; 1.
 DR PROSITE; PS01295; ISPD; 1.
 DR Transferrase; Nucleotidyltransferase; Isoprene biosynthesis;
 KW Complete proteome.
 SQ SEQUENCE 237 AA; 26922 MW; 38AFDEP75BCB706 CRC64;

Query Match 39.6%; Score 474.5; DB 1; Length 237;
 Best Local Similarity 44.4%; Pred. No. 2, 5e-34;
 Matches 99; Conservative 50; Mismatches 67; Indels 7; Gaps 3;

QY 9 AVPAAGVGKMQADRPQYIPLAGKTVIEHTLTRLESDFQKVAVAISVEDPYPELS 68
 Db 13 AVPAAGIGSRMQMDPKQYILHGTILIEHTLSVGLGPILEKILAAVADPYISTCP 72
 QY 69 IAKHPDITAPGGERADSVLSALKALEDA--SEWDVLYVHDAAPPCLTGSDIHQIDFLK 126
 Db 73 ISSNRITISVVGGERKINSVLSGLIVKAV---DWIVHDAVAPCLSTYDLEKLSITIK 128

QY 129 NDVGGIILASHDITKH--VDGDTITATIDRKHWRALTPQMKMLRDALQF--TEGN 185
 DB 129 KNPVGAIIILRPVSDITKSNLQKQKAVTVYVKNMHWALTPQLFQVLELKNCKIKTIQDQ 188
 QY 186 PAVTDASALIELGHEKPKIVEGRPNINIKITRPEDLALAQFYME 228
 DB 189 ISVTDBASALEYCGVNPFLVLGSCRNIKITWPEDLVLANFYLK 231

RESULT 14

ISPD_WIGBR STANDARD; PRT; 229 AA.

AC Q8D23;
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC 2.7.7.60)
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP
 cytidylyltransferase) (MCT).
 OS Wigglesworthia glosiniida brevipalpis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Wigglesworthia.
 NCBI_TaxID=36870;
 [1]
 SEQUENCE FROM N.A.
 RX MEDLINE=22297718; PubMed=12219091;
 RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
 Aksoy S.;
 RT "Genome sequence of the endocellular obligate symbiont of tsetse
 flies, Wigglesworthia glosiniida.";
 RL Nat. Genet. 32:402-407(2002).
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-
 methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-
 phosphate (By similarity).
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB063522; BAC24678.1; --
 DR HAMAP; MF_00108; --; 1.
 DR InterPro; IPR001228; ISPD_synthase.
 DR Pfam; PF01128; ISPD; 1.
 DR TIGRFAMs; TIGR00453; ISPD; 1.
 DR PROSITE; PS01295; ISPD; 1.
 DR TRANSFERS; Nucleotidyltransferase; Isoprene biosynthesis;
 KM Complete proteome.
 SQ SEQUENCE 229 AA; 26240 MW; 9CAlP94717B89FD1 CRC64;

Query Match 38.6%; Score 462.5; DB 1; Length 229;
 Best Local Similarity 44.3%; Pred. No. 2.6e-33;
 Matches 97; Conservative 43; Mismatches 74; Indels 5; Gaps 2;

QY 10 VTPAAGVGKRMQADPRKQTLPLAGKTVIEHTITRLLESDAFOKAVAVSVDPEPELST 69
 DB 10 ITPAAGIGKRMQADPRKQTLPLAGKTVIEHTITRLLESDAFOKAVAVSVDPEPELST 69
 QY 70 AHPGPIITAPGKGRADSVLSALKALIEDIASENDVAVLVDAAAPCLTGSDDIH 129
 DB 70 LKNNKINIVIGKSTESVLSALK-----FVSKVAVVAVHDAAVPCPLHNDKNDKLVKVINI 125
 QY 130 DPVGIIILASHDITKHVDGDTITATIDRKHWRALTPQMKMLRDALQF--TEGNPAV 188
 DB 130 DPVGIIILASHDITKHVDGDTITATIDRKHWRALTPQMKMLRDALQF--TEGNPAV 188

DB 126 SPFGAIIAIPYIDVYKKSYSNGFISHTIKRNMALRALTPOLEFNKILNCKKIITSKBEII 185
 QY 189 TDBASALELGHKPKIVEGRPNINIKITRPEDLALAQFYME 227
 DB 186 TDBASALEKCGYKLVHGRSDINIKITYPEDLNANFFI 224

RESULT 15

ISPD_WIGBR STANDARD; PRT; 236 AA.

AC Q8K9D6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC 2.7.7.60)
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP
 cytidylyltransferase) (MCT).
 OS Buchnera aphidicola (subsp. Schizaphis graminum).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 NCBI_TaxID=98794;
 [1]
 SEQUENCE FROM N.A.
 RX MEDLINE=22084549; PubMed=12089438;
 RA Tamás I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
 Varnegren J.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;
 RT "50 million years of genomic stasis in endosymbiotic bacteria.";
 RL Science 296:2376-2379(2002).
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-
 methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-
 phosphate (By similarity).
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
 diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB014116; AAM67955.1; --
 DR HAMAP; MF_00108; --; 1.
 DR InterPro; IPR001228; ISPD_synthase.
 DR Pfam; PF01128; ISPD; 1.
 DR PIRSF; PIRSF006765; DPCMG_synth; 1.
 DR TIGRFAMs; TIGR00453; ISPD; 1.
 DR PROSITE; PS01295; ISPD; 1.
 DR TRANSFERS; Nucleotidyltransferase; Isoprene biosynthesis;
 KM Complete proteome.
 SQ SEQUENCE 236 AA; 26967 MW; B3CCAlP88740DD25 CRC64;

Query Match 38.2%; Score 457.5; DB 1; Length 236;
 Best Local Similarity 43.3%; Pred. No. 7.5e-33;
 Matches 100; Conservative 42; Mismatches 82; Indels 7; Gaps 3;

QY 2 NPTIOCAVAVPAGVGKRMQADPRKQTLPLAGKTVIEHTITRLLESDAFOKAVAVSVD 61
 DB 6 SPKIAIIVPAGIGKRMQADPRKQTLPLAGKTVIEHTITRLLESDAFOKAVAVSVD 65
 QY 62 PYWPELSIAHPDITTPAGKGRADSVLSALKALIEDIASENDVAVLVDAAAPCLTGSDDIH 121
 DB 66 NFPQKISISSDLRFVSTLGNRHHVSTLGL-----ITTPAKVVIIDAAVAPCLSYDLE 121
 QY 122 LQDPTLKNDDVGGIILASHDITKHVD--GDITITATIDRKHWRALTPQMKMLRDAL 179
 DB 122 NLNITNTKTVGGIILAPVCDITIKSRKKTILHTIIPKQMLHALTPQFPIINLAPCL 181
 QY 180 QR--TEGNPAVTDASALELGHKPKIVEGRPNINIKITRPEDLALAQFYMEQ 229
 DB 180 QR--TEGNPAVTDASALELGHKPKIVEGRPNINIKITRPEDLALAQFYMEQ 229

Db 182 KXIVEDKXNITDPAASHLSTYCGYHPLIVGSYXNIXKITVPEDLIPAEFIKE 232

Search completed: January 29, 2004, 15:50:40
Job time : 5.34994 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:14 ; Search time 17.2185 Seconds
(without alignments)
3461.979 Million cell updates/sec

Title: US-09-941-947a-10
Perfect score: 1199
Sequence: 1 NMPTICWAVVPAAGVGRM.....IKTRPEDLALAQFYMEQQA 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phage:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_virus:.*
16: sp_bacteriophage:.*
17: sp_archaeal:.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	605	50.5	237	16 Q8DC60	Q8DC60 vibrio vuln
2	520.5	43.4	236	16 Q8FEU5	Q8FEU5 escherichia
3	492	41.0	249	16 Q8EBR2	Q8EBR2 shewanella
4	462.5	38.6	229	16 Q8D223	Q8D223 wiglesworth
5	301	25.1	132	2 Q8EYV4	Q8EYV4 klebsiella
6	283.5	23.6	302	10 Q64726	Q64726 arabidopsis
7	272.5	22.7	234	16 Q8DL51	Q8DL51 synecococcus
8	272	22.7	474	2 Q48230	Q48230 haemophilus
9	258	21.5	474	2 Q48154	Q48154 haemophilus
10	256	20.8	238	16 Q8F7A0	Q8F7A0 leptospira
11	249.5	20.7	237	2 Q8G0H4	Q8G0H4 brucella su
12	248.5	20.7	237	2 Q8BK19	Q8BK19 bacillus su
13	248.5	20.5	248	16 Q8FMJ3	Q8FMJ3 corneobacter
14	246	20.5	236	16 Q8Y832	Q8Y832 listeria mo
15	232.5	19.4	246	16 Q8KJ33	Q8KJ33 chlorobium
16	229.5	19.1	240	16 Q8DY07	Q8DY07 streptococcus

17	228.5	19.1	239	16 Q8E4B4	Q8E4B4 streptococcus
18	227	18.9	237	16 Q92CV0	Q92CV0 listeria in
19	224.5	18.7	291	16 Q8G7B2	Q8G7B2 bifidobacter
20	218.5	18.2	238	16 Q8CQ77	Q8CQ77 staphylococcus
21	214.5	17.9	238	16 Q99WM8	Q99WM8 staphylococcus
22	209	17.4	235	16 Q97Q85	Q97Q85 streptococcus
23	209	17.4	235	16 Q8DP12	Q8DP12 streptococcus
24	200.5	16.7	238	16 Q99WX2	Q99WX2 staphylococcus
25	200.5	16.7	240	16 Q8EWC7	Q8EWC7 mycoplasma
26	196.5	16.4	238	16 Q8N1T0	Q8N1T0 staphylococcus
27	171.5	14.3	241	2 Q8GPF2	Q8GPF2 streptococcus
28	170	14.2	324	11 Q8CAE0	Q8CAE0 mus musculus
29	141	11.8	229	17 Q58136	Q58136 pyrococcus
30	133	11.1	203	10 Q8S0P7	Q8S0P7 oryza sativ
31	133	11.1	362	11 Q8BR14	Q8BR14 mus musculus
32	125.5	10.5	246	2 Q3XB85	Q3XB85 amycolatops
33	122	10.2	734	5 Q81273	Q81273 plasmodium
34	118	9.8	458	16 Q9CEP8	Q9CEP8 lactococcus
35	113	9.4	187	11 Q8C934	Q8C934 mus musculus
36	112	9.3	460	16 Q8C5P1	Q8C5P1 bifidobacter
37	109	9.1	492	16 Q9CD44	Q9CD44 mycobacteri
38	109	9.1	513	16 P96382	P96382 mycobacteri
39	108.5	9.0	295	2 Q9X364	Q9X364 bacillus an
40	107.5	9.0	242	16 Q98CM6	Q98CM6 rhizobium 1
41	106.5	8.9	456	16 Q8ZS7	Q8ZS7 yersinia pe
42	106.5	8.9	458	16 Q8CZF5	Q8CZF5 yersinia pe
43	105.5	8.8	291	17 Q9UZ17	Q9UZ17 pyrococcus
44	103.5	8.6	482	16 Q8CJX6	Q8CJX6 streptomyces
45	102	8.5	239	16 Q8FXZ3	Q8FXZ3 brucella su

ALIGNMENTS

RESULT 1

Q8DC60 PRELIMINARY; PRT; 237 AA.
ID Q8DC60;
AC Q8DC60;
DT 01-MAR-2003 (TRENBLREL. 23, Created)
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE 4-diphosphocytidylyl-2-methyl-D-erithritol synthase.
GN VY1582.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
CX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RL Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB016802; AAO10005.1; -.
KW Complete proteome.
SQ SEQUENCE 237 AA; 26005 MW; 35D0952D12B47ECD CRC64;

Query Match 50.5%; Score 605; DB 16; length 237;
Best Local Similarity 55.4%; Prod. No. 9.6e-44;
Matches 124; Conservative 31; Mismatches 67; Indels 2; Gaps 2;
QY 9 AVYPAAGVGRMADREKQYVPLAGKTVIEHTLRLLESAPQVAVVAVSDEPYNELS 68
DB 10 AVYPAAGVGRMADREKQYVPLAGKTVIEHTLRLLESAPQVAVVAVSDEPYNELS 69
QY 69 IAKPDIITAPGKERADSVLSALKALEDIASERDWTAVDAPPCULGSDIHQIDTLK 128
DB 70 IAKPDIITAPGKERADSVLSALKALEDIASERDWTAVDAPPCULGSDIHQIDTLK 129
QY 129 NDPVGGIILASSHTTLKAVUG-DITITIRKRVKVALTQMPKRYGLRDAL-QRTGNP 186
DB 130 SHETGILATPVRDITMKRANAGQWIDHTVRNALMHALTFQMPKRYGLRDAL-QRTGNP 189

QY 187 AVTDEASALELGHKPKIVEGRPNKIKTRPEDLALAQFYMQ 230
 DB 130 ALTDEASALEMGRBEPALVQSCSSNKVTPEDLALAEYLSRE 233

RESULT 2

Q8E8T5 PRELIMINARY; PRT; 236 AA.

AC Q8E8T5; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60).
 GN YGAP OR C3314.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;

SEQUENCE FROM N.A.
 RC STRAIN-06.H1 / CPT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasco D., Buckles B.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner P.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AB016765; AAN01763.1; -;
 KW Nucleotidyltransferase; Transferase; Complete proteome.

Query Match 43.4%; Score 520.5; DB 16; Length 236;
 Best Local Similarity 49.8%; Pred. No. 1.6e-36;

Matches 111; Conservative 32; Mismatches 71; Indels 9; Gaps 4;

QY 9 AVVPAAGVKRMQADRPKQYLPPLAGKTVIEHTLRLLESDAFOKVAVAISVEDPYMPELS 68
 DB 10 AVVPAAGFRMRQTECPKQYLSIGNQTLLESHVALLHPRKRVVAISPEDSRPAQLP 69
 QY 69 IAKHPDITAPGKERADSVLSALKLEDIASENDVVLVHDAARPCLTGSDIHQIDTLK 128
 DB 70 IANHPRIIVVDGEBERADSVLAKLKA---AGDAQWVIVHDAARPCLTGSDIHQIDTLK 125
 QY 129 NDPVGGIILASSHTLKHVG--DTTATIDRKHWRAALTPQMFYGMRLDALQ--TEG 184
 DB 126 TERTGIIAAPVDIMKRAEPGRKNAIHTVDKNGIMHAIPTQFFRELLHDCITLALNG 185
 QY 185 NPAVTDEASALELGHKPKIVEGRPNKIKTRPEDLALAQFYM 227
 DB 186 ALTDEASALEYCGFHPQVLEGRADNIKTRPEDLALAEFYL 227

RESULT 3

Q8E8R2 PRELIMINARY; PRT; 249 AA.

AC Q8E8R2; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE 4-diphosphocytidylyl-2C-methyl-D-erythritol synthase.
 GN ISPD OR S03438.

OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;

SEQUENCE FROM N.A.
 RC STRAIN=NR-1;
 RX MEDLINE=22297686; PubMed=12368813;

RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos B.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seehardt R., Ward N., Methe B., Clayton R.A.,
 RA Meyer R., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.U., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umeyam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouir H., Gill J., Utterback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealeon K.H., Fraser C.M.;
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 Shewanella oneidensis.";
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL; AB015780; AAN56435.1; -;
 RX TIGR; S03438; -;
 KW Complete proteome.

Query Match 41.0%; Score 492; DB 16; Length 249;
 Best Local Similarity 48.0%; Pred. No. 4.7e-34;
 Matches 106; Conservative 33; Mismatches 76; Indels 6; Gaps 3;

QY 9 AVVPAAGVKRMQADRPKQYLPPLAGKTVIEHTLRLLESDAFOKVAVAISVEDPYMPELS 68
 DB 28 AVVPAAGISRMWGRKQYLPPLGOSILAHLDKLSHPLISQVYALPEDADFYALP 87
 QY 69 IAKHPDITAPGKERADSVLSALKLEDIASENDVVLVHDAARPCLTGSDIHQIDTLK 128
 DB 88 QAKHPKIKTVYGGSERANSVLAAL---DKAPNSMALVHDAARPCLMASDIDDLTSRV 143
 QY 129 NDPVGGIILASSHTLKHVG--DTTATIDRKHWRAALTPQMFYGMRLDALQ--RTG 186
 DB 144 QFPGAILAMPVRRTMKRASLSGINSIVCRDWMHAIPTQLPFTSLRLHLOALNAGA 203
 QY 187 AVTDEASALELGHKPKIVEGRPNKIKTRPEDLALAQFYM 227
 DB 204 VTTDEASAMEMAGISPLVAGRADNIKTRPEDLALAEFL 244

RESULT 4

Q8D223 PRELIMINARY; PRT; 229 AA.

AC Q8D223; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE YGP protein.
 GN YGP.

OS Wigglesworthia brevipalpis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Wigglesworthia.
 OX NCBI_TaxID=164609;

SEQUENCE FROM N.A.
 RC MEDLINE=22297718; PubMed=12219091;
 RA Akman L., Yamashta A., Watanabe H., Oshima K., Shiba T., Hattori M.,
 RA Akao S.;
 RT "Genome sequence of the endocellular obligate symbiont of tsetse
 fly flies, Wigglesworthia glossinidia.";
 RL Nat. Genet. 32:402-407(2002).
 RX EMBL; AB063522; BAC24678.1; -;
 KW Complete proteome.

Query Match 38.6%; Score 462.5; DB 16; Length 229;
 Best Local Similarity 44.3%; Pred. No. 1.4e-31;

Matches 97; Conservative 43; Mismatches 74; Indels 5; Gaps 2;

QY 10 VVPAAGVKRMQADRPKQYLPPLAGKTVIEHTLRLLESDAFOKVAVAISVEDPYMPELS 69
 DB 10 IIPPAAGIKRMQYCPKQYLIKINKTLEHSISLFDIKYVKLLIANKKDYFNFSLI 69
 QY 70 AKHPDITAPGKERADSVLSALKLEDIASENDVVLVHDAARPCLTGSDIHQIDTLK 129
 DB 70 LKNNKINIVIGKSRISVLSLK---FVSKVDWVIVHDAARPCLTGSDIHQIDTLK 125

QY 130 DPGGILATSSHTLKHVQSDTITATDRKHWALTPQWKYGMPLDNLQ-RTBGNPAV 188
 DB 126 SPFGALIAPIYIVKVSYNFISHTIKRKLMPALPOLFNKILNCKIITSGEII 185
 QY 189 TDEASALELGHKPKIVEGRPDNIKITRPEDLALAOFPYM 227
 DB 186 TDESSALEKCGYLNLVHGSNDNIKITYPEDLNPNANFI 224

RESULT 5
 QY09EY4 PRELIMINARY; PRT; 132 AA.
 ID Q9EY4; AC Q9EY4;
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE YBP (Fragment).
 GN YBP.
 OS Klebsiella aerogenes.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 OX NCBI_TaxID=28451;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N70;
 RX MEDLINE=20566700; PubMed=11114933;
 RA Kolko M.M., Kapetanovich L.A., Lawrence J.G.;
 RT "Alternative pathways for streptococcus synthesis in Klebsiella
 aerogenes.";
 RT Bacteriol. 183:328-335(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N70;
 RA Seifstein T.A., Lawrence J.G.;
 RT "Methionine recycling in Klebsiella aerogenes.";
 RT Bacteriol. 0:0-0(2001).
 DR EMBL; AF308468; AAC42460.1; -
 DR HSSP; Q46893; 1152.
 DR InterPro; IPR001228; ISPD_synthase.
 DR Pfam; PF01128; ISPD; 1.
 DR PROSITE; PS01295; ISPD; 1.
 DR NON TER 132 132
 SQ SEQUENCE 132 AA; 14033 MW; 69F919A5084B2AFB CRC64;

Query Match 25.1%; Score 301; DB 2; Length 132;
 Best Local Similarity 48.8%; Pred. No. 4.1e-18;
 Matches 62; Conservative 18; Mismatches 43; Indels 4; Gaps 1;

QY 9 AVTPAGVGRKQADRPKQYLPLAGKTVIEHTITRLLESDAFQKVAIVASVEDPYWPELS 68
 DB 10 AVTPAGVGRKQADRPKQYLPLAGKTVIEHTITRLLESDAFQKVAIVASVEDPYWPELS 69
 QY IAGHPDITAPGKGRADSVLSALKALIEDIASNDVVLVHDAARPLCTSGSDIHLAQIDTLK 128
 DB 70 IAGHPDITAPGKGRADSVLSALKALIEDIASNDVVLVHDAARPLCTSGSDIHLAQIDTLK 125
 QY 129 NDVPGGI 135
 DB 126 TSVVGGI 132

RESULT 6
 QY064726 PRELIMINARY; PRT; 302 AA.
 ID Q64726; AC Q64726;
 DT 01-AUG-1998 (TRENBLrel. 07, Created)
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Putative sugar nucleotide phosphorylase
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase)
 DE (2-C-methyl-D-erythritol 4-phosphate cytidyltransferase).
 GN ATG302500 OR ISPD OR ATWEPCT OR ATG202500/T8K22.20.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustersids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Lin X., Kaul S., Tomlin C.D., Benito M.-I., Creeasy T.H., Haas B.J.,
 Wu D., Walz R., Rongling C.M., Koo H., Fujii C.Y., Uterback T.R.,
 Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome 2 BAC T8K22 genomic sequence.";
 RT Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20300921; PubMed=10841550;
 RA Rohdich F., Wungstentweckul J., Eisenreich W., Richter G.,
 Schuhr C.A., Hecht S., Zenk M.H., Bacher A.;
 RT "Biosynthesis of terpenoids. 4-diphosphocytidylyl-2C-methyl-D-erythritol
 synthase of Arabidopsis thaliana.";
 RT Proc. Natl. Acad. Sci. U.S.A. 97:6451-6456(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RA Okada K., Kawade H., Kuzuyama T., Takagi M., Seto H., Kamiya Y.;
 RT "2-C-methyl-D-erythritol 4-phosphate cytidyltransferase.";
 RT Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
 Nakajima M., Enji A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
 Hayashizaki Y., Shinozaki K.;
 RT "Arabidopsis thaliana full-length cDNA.";
 RT Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AC004136; AAC18936.2; -
 DR EMBL; AF230737; AAF61714.1; -
 DR EMBL; AB037876; BAB21592.1; -
 DR EMBL; AK118110; BAC42737.1; -
 DR HSSP; Q46893; 1152.
 DR InterPro; IPR001228; ISPD_synthase.
 DR Pfam; PF01128; ISPD; 1.
 DR TIGRFAWS; TIGR00453; ISPD; 1.
 DR PROSITE; PS01295; ISPD; 1.
 DR TRANSFERASE.
 SQ SEQUENCE 302 AA; 33937 MW; 7881DC5C8C37B06 CRC64;

Query Match 23.6%; Score 283.5; DB 10; Length 302;
 Best Local Similarity 34.2%; Pred. No. 4e-16;
 Matches 80; Conservative 46; Mismatches 83; Indels 25; Gaps 10;

QY 10 VTPAGVGRKQADRPKQYLPLAGKTVIEHTITRLLESDAFQKVAIVASVEDPYWPELS 63
 DB 82 ILLAGGQGRKQADRPKQYLPLAGKTVIEHTITRLLESDAFQKVAIVASVEDPYWPELS 135
 QY 64 WPELSIAKHDPDITAPGKGRADSVLSALKALIEDIASNDVVLVHDAARPLCTSGSDIHL 122
 DB 136 IFEYTESIDVDFALPAGKGRQDSVYSGLOEI-DVNSE--LVCHDSARPLVWTEVE- 191
 QY 123 QIDTLKNDPVGG--ILLASHTLKHVQSDTITATDRKHWALTPQWKYGMPLDNLQ-RTBGNPAV 178
 DB 192 --KVLKGSNAVGAIVAGPAKATIKVYVSDSLVYKTDRLKTLMEQIQVIRPELTKGKF 249
 QY 179 -LQRTGKPAVTDASALELGHKPKIVEGRPDNIKITRPEDLALAOFPYM 231
 DB 250 ELVNSEG-LEVTLDVSVIVELKHPVYVSGSYTNIKVYTPDDLLLAERILSEDS 302

RESULT 7
 QY08D191 PRELIMINARY; PRT; 234 AA.
 ID Q8D191; AC Q8D191;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE 4-diphosphocytidylyl-2C-methyl-D-erythritol synthase.
 OS TIR0605.
 GN Synecococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=22225144; PubMed=12240634;
 RA Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Itiguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsuno M., Matsuno C., Nakazaki N.,
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynechococcus elongatus BP-1."
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AP005371; BAC08157.1; -.
 KW Complete proteome.
 SQ SEQUENCE 234 AA; 25592 MW; EEBFDBEB32BC18E CRC64;

Query Match 32.7%; Score 272.5; DB 16; Length 234;
 Best Local Similarity 35.1%; Pred. No. 2.5e-15;
 Matches 80; Conservative 38; Mismatches 97; Indels 13; Gaps 7;
 QY 10 VVPAAGGKRMQADPRQYLPAGKTVLEHTLRLLESAPQKVAVAISVED-PYV--- 64
 DB 4 LIPAGMGKRMGSHNKLRLQLGKPLAWTLAAVAALIMWIGVIGQPEDFPIEMALL 63
 QY 65 PELSLAKHEDITTAGGKERADSVLSALKALADIASENDVTVHDAAPCLTGSIDHQT 124
 DB 64 EDLNRQPVHLLT-GEETQASVPHGLAPKTEQ--VLIHGGR-CLATPPLINRC 117
 QY 125 DELKNDPVGGILALSHDITLKVDCG-TTATIDRKHWRALTPOMFKYGMRLDQRT 183
 DB 118 AGALGTAGLILAAVPRKOTIKIVNEGAVVCTPREDSLMAQTPGFRVEPLRIHMAV 177
 QY 184 GNP-AVTDASALELIGKPKIVBGPDKIKITRPEDLALQFYEQ 230
 DB 178 AKGWEVTDALFERLGVAVHIVLGEETLKITTPSDIPLERILQHR 225

RESULT 8

Q48230 PRELIMINARY; PRT; 474 AA.
 ID Q48230
 AC Q48230;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Bifunctional ribulose 5-phosphate reductase/CDP-ribitol
 DE pyrophosphorylase.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RM135;
 RX MEDLINE=95272382; PubMed=7752865;
 RA van Eldere J., Brophy L., Loynds B., Celis P., Kroil J.S., Moxon E.R.,
 RA Hancock I., Carman S.;
 RT "Region II of Haemophilus influenzae type b capsulation locus is
 RT involved in serotype-specific polysaccharide synthesis."
 RL Mol. Microbiol. 15:107-118(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RM135;
 RX MEDLINE=99194706; PubMed=10094675;
 RA Follens A., Veiga-da-Cunha M., Merckx R., an Schaftingen E.,
 RA van Eldere J.;
 RT "Accl of Haemophilus influenzae type a Capsulation Locus Region II
 RT encodes a bifunctional ribulose 5-phosphate reductase-CDP-ribitol

RT pyrophosphorylase."
 RL J. Bacteriol. 181:2001-2007(1999).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASIS/REDUCTASES
 CC (SPR) FAMILY.
 DR EMBL; X78559; CAA55303.1; -.
 DR InterPro; IPR002198; ADH_short.
 DR InterPro; IPR001228; ISPd synthase.
 DR Pfam; PF00106; adh_short; 1.
 DR Pfam; PF01128; ISPd; 1.
 DR PROSITE; PS00061; ADH_SHORT; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 474 AA; 52467 MW; DC42A2B245318A95 CRC64;

Query Match 22.7%; Score 272; DB 2; Length 474;
 Best Local Similarity 34.6%; Pred. No. 7.3e-15;
 Matches 81; Conservative 34; Mismatches 85; Indels 34; Gaps 9;
 QY 10 VVPAAGGKRMQADPRQYLPAGKTVLEHTLRLLESAPQKVAVAISVED-PYV 63
 DB 9 IILAGVSRMGICYPQPSFKLAKTALFQEHKEIDELIIVSERTSYRIED-I 67
 QY 64 MPELSIAKHEDITTAGGKERADSVLSALKALADIASENDVTVHDAAPCLTGSIDH 114
 DB 68 VSKLDPSKVNRIIR-GEKESDSTLSAITPLAD-BEENYKLIHDVAPPLATEIIS 124
 QY 115 LTGSDIHQIDITKNDPVGGILALSHDITLKVDCG-TTATIDRKHWRALTPOMFKY 172
 DB 125 IAKDKVNAVD-----VAIPAVDTIVHVNNDQEIILKIPRAEYVGQGPQAFKL 174
 QY 173 GMLRDALQ-RTGKNPAVTDASAL--RLHGKPKIVBGPDKIKITRPEDLALQ 223
 DB 175 GTLAKAVDIYQGGIEGTCCSTVLKTLPEERVGIVSGSETNIKLTSPVDLFIA 228

RESULT 9

Q48154 PRELIMINARY; PRT; 474 AA.
 ID Q48154
 AC Q48154;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Bifunctional ribulose 5-phosphate reductase/CDP-ribitol
 DE pyrophosphorylase.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 9006;
 RA Celis P., Moxon E., Byssen H., van Eldere J.;
 RT "Genetic analysis of the Region II of the Haemophilus influenzae
 RT serotype a capsulation locus."
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 9006;
 RA Celis P.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 9006;
 RX MEDLINE=99194706; PubMed=10094675;
 RA Follens A., Veiga-da-Cunha M., Merckx R., an Schaftingen E.,
 RA van Eldere J.;
 RT "Accl of Haemophilus influenzae type a Capsulation Locus Region II
 RT encodes a bifunctional ribulose 5-phosphate reductase-CDP-ribitol
 RT pyrophosphorylase."
 RL J. Bacteriol. 181:2001-2007(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 9006;
 RA Follens A.;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 ; Search time 19.4547 Seconds
(without alignments)
3044.503 Million cell updates/sec

Title: US-09-941-947a-12

Perfect score: 1516

Sequence: 1 MDVYAGGGERMPAPAKNTLM.....LVFLAKLNGSALYKLEQG 285

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 789580 seqs, 207824079 residues 789580

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1516	100.0	285	US-09-934-903-8	Sequence 8, Appl1
2	1516	100.0	285	US-09-934-868-74	Sequence 74, Appl1
3	1516	100.0	285	US-09-941-947a-12	Sequence 12, Appl1
4	638	42.1	279	US-10-369-493-13766	Sequence 13766, A
5	620	40.9	279	US-10-369-493-201	Sequence 201, App
6	616	40.6	283	US-10-369-493-802	Sequence 802, App
7	584.5	38.5	284	US-10-369-493-21079	Sequence 21079, A
8	584	38.4	294	US-10-369-493-15324	Sequence 15324, A
9	582	38.4	276	US-10-369-493-15690	Sequence 15690, A
10	582	38.4	276	US-10-369-493-16082	Sequence 16082, A
11	580.5	38.3	298	US-10-369-493-17751	Sequence 17751, A
12	572	37.7	270	US-10-369-493-9184	Sequence 9184, Ap
13	572	37.7	270	US-10-369-493-9359	Sequence 9359, Ap
14	556	36.7	213	US-10-369-493-19675	Sequence 19675, A
15	533	35.2	233	US-10-369-493-8419	Sequence 8419, Ap

16	407	26.8	284	US-10-369-493-9956	Sequence 9956, Ap
17	341	22.5	289	US-10-369-493-16541	Sequence 16541, A
18	327	21.6	269	US-10-369-493-7851	Sequence 7851, Ap
19	324	21.4	289	US-10-369-493-23046	Sequence 23046, A
20	320	21.1	257	US-10-369-493-9660	Sequence 9660, Ap
21	317	20.9	287	US-10-369-493-17097	Sequence 17097, A
22	292	19.3	316	US-10-369-493-20265	Sequence 20265, A
23	286.5	18.9	325	US-10-369-493-12283	Sequence 12283, A
24	281	18.5	317	US-10-369-493-11123	Sequence 11123, A
25	278	18.3	324	US-10-369-493-11123	Sequence 11123, A
26	274	18.1	292	US-10-369-493-11764	Sequence 11764, A
27	273.5	18.0	261	US-10-369-493-11764	Sequence 11764, A
28	273.5	18.0	261	US-10-369-493-14419	Sequence 14419, A
29	273.5	18.0	261	US-10-369-493-14670	Sequence 14670, A
30	273.5	18.0	261	US-10-369-493-15150	Sequence 15150, A
31	270.5	17.8	305	US-10-369-493-15150	Sequence 15150, A
32	267	17.6	315	US-09-712-363-186	Sequence 2607, Ap
33	254	16.8	240	US-10-369-493-17829	Sequence 17829, Ap
34	250.5	16.5	268	US-10-369-493-65	Sequence 65, Appl
35	239	15.8	295	US-10-369-493-16874	Sequence 16874, A
36	233.5	15.4	271	US-10-369-493-3024	Sequence 3024, Ap
37	216	14.2	311	US-09-738-626-4515	Sequence 4515, Ap
38	203	13.4	245	US-10-369-493-10645	Sequence 10645, A
39	178	11.7	141	US-10-369-493-1236	Sequence 1236, Ap
40	160	10.6	99	US-10-264-213-208	Sequence 208, App
41	127	8.4	238	US-09-882-227-358	Sequence 358, App
42	125	8.2	314	US-09-815-242-10963	Sequence 10963, A
43	116	7.7	575	US-10-104-047-2815	Sequence 2815, Ap
44	116	7.7	813	US-10-108-2604-3901	Sequence 3901, Ap
45	116	7.7	990	US-10-094-749-2568	Sequence 2568, Ap

ALIGNMENTS

RESULT 1
US-09-934-903-8
Sequence 8, Application US/09934903
Patent No. US20020102690A1
GENERAL INFORMATION:
APPLICANT: Koffas, Mattheos
APPLICANT: Odem, J. Martin
APPLICANT: Schenzle, Andreas J.
APPLICANT: No. US20020102690A1, Kelley C.
APPLICANT: Tomb, Jean-Francois
APPLICANT: Rouviere, Pierre
APPLICANT: Piccataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
FILE REFERENCE: C1646 US NA
CURRENT APPLICATION NUMBER: US/09/934,903
CURRENT FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/229,907
PRIOR FILING DATE: September 1, 2001
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 8
LENGTH: 285
TYPE: PRT
ORGANISM: Methylobionas 16a
FEATURES:
OTHER INFORMATION: Amino acid sequences encoded by ORF4
US-09-934-903-8

Query Match 100.0%; Score 1516; DB 10; Length 285;
Best Local Similarity 100.0%; Pred. No. 5.7e-156;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDVYAGGGERMPAPAKNTLMRTGRRPGYHLOTFQMDLCMDLTPHPVDGRTLR 60
DB 1 MDVYAGGGERMPAPAKNTLMRTGRRPGYHLOTFQMDLCMDLTPHPVDGRTLR 60
QY 61 NPISVPEQDDLTFRANLLKSHTCVRCVCIDIERNLPMGSGGLGSGSSDAATTVLVLR 120

```

Db      61 NPISGVPEODDLTVRAANLTKSHTCVAGVCIDIEKRLPMGGGLGGSSDAATLVNLR 120
      121 LMGGLSKRELMDLGLRLGADVVFVFGCSAMGEGVSEDLOATILPEQMFVILKPDCHVN 180
      121 LMGGLSKRELMDLGLRLGADVVFVFGCSAMGEGVSEDLOATILPEQMFVILKPDCHVN 180
      181 TGEIFSAENLTRNSAVVTVMSDFLAGDNNDCEVVCCKYRPVKADIDALLCYAARLTGT 240
      181 TGEIFSAENLTRNSAVVTVMSDFLAGDNNDCEVVCCKYRPVKADIDALLCYAARLTGT 240
      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEEG 285
      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEEG 285

```

RESULT 2

US-09-934-868-74
Sequence 74, Application US/09934868
Patent No. US20020137190A1

GENERAL INFORMATION:

APPLICANT: Koffas, Mattheos
APPLICANT: Odom, James M.
APPLICANT: Schenck, Andreas J.
TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN
FILE REFERENCE: CL1596 US NA
CURRENT APPLICATION NUMBER: US/09/934,868
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/229,858
NUMBER OF SEQ ID NOS: 81
SOFTWARE: Microsoft Office 97
SEQ ID NO 74
LENGTH: 285
TYPE: PRT
ORGANISM: Methylobionas 16a

FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by ISPE
US-09-934-868-74

Query Match 100.0%; Score 1516; DB 10; Length 285;
Best Local Similarity 100.0%; Pred. No. 5.7e-156;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

      1 MDYAAGWGERPAPAKIMLMLRITGRPDGYHLQTVFQMLDCDMLTFHPVDGRTLR 60
      1 MDYAAGWGERPAPAKIMLMLRITGRPDGYHLQTVFQMLDCDMLTFHPVDGRTLR 60
      61 NPISGVPEODDLTVRAANLTKSHTCVAGVCIDIEKRLPMGGGLGGSSDAATLVNLR 120
      61 NPISGVPEODDLTVRAANLTKSHTCVAGVCIDIEKRLPMGGGLGGSSDAATLVNLR 120
      121 LMGGLSKRELMDLGLRLGADVVFVFGCSAMGEGVSEDLOATILPEQMFVILKPDCHVN 180
      121 LMGGLSKRELMDLGLRLGADVVFVFGCSAMGEGVSEDLOATILPEQMFVILKPDCHVN 180
      181 TGEIFSAENLTRNSAVVTVMSDFLAGDNNDCEVVCCKYRPVKADIDALLCYAARLTGT 240
      181 TGEIFSAENLTRNSAVVTVMSDFLAGDNNDCEVVCCKYRPVKADIDALLCYAARLTGT 240
      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEEG 285
      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEEG 285

```

RESULT 3

US-09-941-947A-12
Sequence 12, Application US/09941947A
Patent No. US20030003528A1

GENERAL INFORMATION:

APPLICANT: Brzostowicz, Patricia C.
APPLICANT: Cheng, Qiong
APPLICANT: Dicosimo, Deana J.

```

      APPLICANT: Koffas, Mattheos
      APPLICANT: Miller, Edward S. Jr.
      APPLICANT: Odom, J. Martin
      APPLICANT: Picotaggio, Steve
      APPLICANT: Rouviere, Pierre E.
      TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
      FILE REFERENCE: CL1903 US NA
      CURRENT APPLICATION NUMBER: US/09/941,947A
      PRIOR FILING DATE: 2001-09-01
      PRIOR APPLICATION NUMBER: 60/229,907
      PRIOR FILING DATE: 2000-09-01
      PRIOR APPLICATION NUMBER: 60/229,858
      NUMBER OF SEQ ID NOS: 60
      SOFTWARE: Microsoft Office 97
      SEQ ID NO 12
      LENGTH: 285
      TYPE: PRT
      ORGANISM: Methylobionas 16a
      US-09-941-947A-12

```

Query Match 100.0%; Score 1516; DB 11; Length 285;
Best Local Similarity 100.0%; Pred. No. 5.7e-156;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

      1 MDYAAGWGERPAPAKIMLMLRITGRPDGYHLQTVFQMLDCDMLTFHPVDGRTLR 60
      1 MDYAAGWGERPAPAKIMLMLRITGRPDGYHLQTVFQMLDCDMLTFHPVDGRTLR 60
      61 NPISGVPEODDLTVRAANLTKSHTCVAGVCIDIEKRLPMGGGLGGSSDAATLVNLR 120
      61 NPISGVPEODDLTVRAANLTKSHTCVAGVCIDIEKRLPMGGGLGGSSDAATLVNLR 120
      121 LMGGLSKRELMDLGLRLGADVVFVFGCSAMGEGVSEDLOATILPEQMFVILKPDCHVN 180
      121 LMGGLSKRELMDLGLRLGADVVFVFGCSAMGEGVSEDLOATILPEQMFVILKPDCHVN 180
      181 TGEIFSAENLTRNSAVVTVMSDFLAGDNNDCEVVCCKYRPVKADIDALLCYAARLTGT 240
      181 TGEIFSAENLTRNSAVVTVMSDFLAGDNNDCEVVCCKYRPVKADIDALLCYAARLTGT 240
      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEEG 285
      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEEG 285

```

RESULT 4

US-10-369-493-13766
Sequence 13766, Application US/10369493
Patent No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10 (52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 13766
LENGTH: 279
TYPE: PRT
ORGANISM: Pseudomonas fluorescens
US-10-369-493-13766

Query Match 42.1%; Score 638; DB 12; Length 279;
Best Local Similarity 49.6%; Pred. No. 1.2e-60;
Matches 135; Conservative 41; Mismatches 96; Indels 0; Gaps 0;

QY 12 PAAKNTLMRIIGRRPDGTHLLQTFQFMQDLCMTLPHRVPDGRVTLRNPLSGVPEQDD 71

DQ 5 PSPAKNLMLHLIGRREDGTHLQTLFQFLDIDODETTFAVRDQVIRLHLTEPDGPHPSN 64

QY 72 LTPRANLLKSHTGCVARGVCIDIERNLPMGGGIGGGSSDAATTVILNRLMGLSKREL 131

DQ 65 LTPRANQDIOBOSACSGIDITWIDKLLPMGGGIGGGSSNAATTVLLGLNHLRLGMDSDRL 124

QY 132 MDVGLRIGADVPRFVFGCSAMGEGVSEDLCATLTPRPFVILIKPDCHNVTGIFSGAEMLT 199

DQ 125 AALGLTLGADVPRFVARGHAAPAGVGEKLTLPBPAPRWVTVLPQVSSTABIESDPLLT 188

QY 192 RNSAVVTTMSDPLAGDNRNDQSEVYCKLYRPVKDAIDALICVAEARLTGTGA CVFAQFCNK 251

DQ 185 RNPPLPKVPRVPGNSRNDCLPVFSRRYPRVPRNALNLGNFTEAKLTGTGSCVFGGFPSSK 244

QY 252 EDASALLEGKDRWLFLAKGLNQSLYYKLE 283

DQ 245 AEAADKVSALLTETLTGFAVAKGSSVSLNHRKIQ 276

```

RESULT 5
US-10-369-493-201
; Sequence 201, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; PITS REFERENCES: 38-10 (52052) B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 201
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
; US-10-369-493-201

```

Query Match	40.9%;	Score 620;	DB 182;	Length 279;
Best Local Similarity	48.0%;	Pred. No. 1; DE 58;		
Matches 122;	Conservative 36;	Mismatches 101;	Indels 6;	Gaps 2
QY	11	WRPAPATLNTMLKTRGRPDPTGHTLTQVMTDLCLMTLTFHWDDGRVTLRNPISGVPEOD	70	
Db	5	WSPAPATLNTFLVITGRRSDGYHRLQTLFQFLPYGGEITITTRQDNQIRLLPLPVKGVNDD	64	
QY	71	DLTVRAANLTKSH-----TGC-VRGVCVIDIEKNLPMGGSLGGSSDAATTLVVLNRLMWL	124	
Db	65	NILVVRARLLQDAHEKQSGTQGYRRADHIHKRLPYGGSLGGSSNAAITVLALNHYMQT	124	
QY	125	GISKPELMDLGLRLGADVPLVFGCSAMGEGVSEDLQAITLPQGMPTIKPDQVNTGRT	184	
Db	125	NISDDELQGLGVTLLGADVPVFKVGGAAAEAGEGEQLQPSASEEEMFVFAHVGHIEIPPKI	184	
QY	185	FAEENLITRNSAVVTWMSDFAGDNRNDCSEVVCYKLRPVKDAIDALLCAEARLITGTGACV	244	
Db	185	FTDPDELKRRNSPIRTLPALLQAPPNKDCPIARKRPREVEGLSMLLLETTSRLTGACV	244	
QY	245	FAQFCKEDAEASALEGLDNRMLVFLAKKLNSALT	279	
Db	245	FGFESFSPASARKVLNQAPRMQGFVARKVNSLPH	279	

RESULT 6
US-10-369-493-802

? Sequence 802, Application US/10369493
 ? Publication No. US20030233675A1
 ? GENERAL INFORMATION:
 ? APPLICANT: Cao, Yongwei
 ? APPLICANT: Hinkle, Gregory J.
 ? APPLICANT: Slater, Steven C.
 ? APPLICANT: Goldman, Barry S.
 ? APPLICANT: Chen, Xianfeng
 ? TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ? TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ? FILE REFERENCE: 38-10 (52052) B
 ? CURRENT APPLICATION NUMBER: US/10/369,493
 ? CURRENT FILING DATE: 2003-02-28
 ? PRIOR APPLICATION NUMBER: US 60/360,039
 ? PRIOR FILING DATE: 2002-02-21
 ? NUMBER OF SEQ. ID NOS: 47374
 ? SEQ. ID NO 802
 ? LENGTH: 283
 ? TYPE: PRT
 ? ORGANISM: Escherichia coli
 ? US-10-369-493-802

[illegible]

```

1      RESULT 7
2      US-10-369-493-21079
3      ; Sequence 21079, Application US/10369493
4      ; Publication No. US20030233675A1
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Cao, Yongwei
7      ; APPLICANT: Hinkle, Gregory J.
8      ; APPLICANT: Slater, Steven C.
9      ; APPLICANT: Goldman, Barry S.
10     ; APPLICANT: Chen, Xianfeng
11     ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
12     ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
13     ; FILE REFERENCE: 38-10(52052)B
14     ; CURRENT APPLICATION NUMBER: US/10/369,493
15     ; CURRENT FILING DATE: 2003-02-28
16     ; PRIOR APPLICATION NUMBER: US 60/360,039
17     ; PRIOR FILING DATE: 2002-02-21
18     ; NUMBER OF SEQ ID NOS: 47374
19     ; SEQ ID NO 21079
20     ; LENGTH: 280
21     ; TYPE: prt
22     ; ORGANISM: Xenorhabdus nematophilus
23     ; S-10-369-493-21079

```

Query Match	38.6%;	Score 584.5;	DB 12;	Length 280;
Best Local Similarity	45.3%;	Pred. No. 7.8e-55;		

Matches 125; Conservative 42; Mismatches 102; Indels 7; Gaps 2;

QY 11 WPAPAKLMLRTGRRPDGTHLLQTVFQMLDCLMTFHPVDDGRV-TLRNPISGVPEQ 70
 Db 5 WSPAPAKLMLRTGRRPDGTHLLQTVFQMLDCLMTFHPVDDGRV-TLRNPISGVPEQ 64
 QY 71 DLTVRAANLTKSH-----TGCVR-GVCIDIEKNLPMGGGLGGSSDAATTVLNLMLG 123
 Db 65 NLIVRAARLQDHLQNNKNGTGLHADIHDKLPMGGGLGGSSDAATTVLNLMLG 124
 QY 124 LGSIKRELMDGLRIGADVPEVFGCSAMGSEVEDLQATLPEQWFIKKPDCHVNTGE 183
 Db 125 ANLSDEDLAQGISIGADVPEVFGKHAFAEGIGEKLRPASPEEKPLVHPGISISTPL 184
 QY 184 LFSKENTRSAAVTMSDFLAGNNDCESEVCKLYRPVDAIDALLCYAERLTGTGAC 243
 Db 185 LFTDPELKRNSPKLSLPAQLQAPFTNDCESIARRFRKVKQLLSWLEVPASLTGTGAC 244
 QY 244 VFAQCKEDAESALBGLKDRWLVFLAKLNOSALY 279
 Db 245 VFGFSEASARKVLNDSPEWMOGFVARGVNIISPLH 280

RESULT 8
 US-10-369-493-15324
 ; Sequence 15324, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 15324
 ; LENGTH: 294
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas campestris
 ; US-10-369-493-15324

Query Match 38.5%; Score 584; DB 12; Length 294;
 Best Local Similarity 45.8%; Pred. No. 9.5e-55;
 Matches 126; Conservative 42; Mismatches 105; Indels 2; Gaps 2;

QY 11 WPAPAKLMLRTGRRPDGTHLLQTVFQMLDCLMTFHPVDDGRV-TLRNPISGVPEQ 69
 Db 17 WSPAPAKLMLRTGRRPDGTHLLQTVFQMLDCLMTFHPVDDGRV-TLRNPISGVPEQ 76
 QY 70 DLTVRAANLTKSHGCVGVCIDIEKNLPMGGGLGGSSDAATTVLNLMLG 123
 Db 77 DDLWRAARALQIHAGTALGAEIRYDKRIPAGGFGGSSDAATTVLNLMLG 136
 QY 130 ELMDELRLGADVPEVFGCSAMGSEVEDLQATLPEQWFIKKPDCHVNTGE 189
 Db 137 TLAEGLRLGADVPEVFGKHAFAEGIGEKLRPASPEEKPLVHPGISISTPL 184
 QY 190 LTRNSAAVTMSDFLAGNNDCESEVCKLYRPVDAI-DALLCYAERLTGTGAC 248
 Db 197 LTRDAAPAKIADPAGSSLLDNAPFVLRREPAIEAVFQALSRIGTRLTGSGSCFVEF 256
 QY 249 CCKEDAESALBGLKDRWLVFLAKLNOSALY 283
 Db 257 ATRAAAEQAMHLPENLRAMWVEGAHSPLLDALD 291

RESULT 9

US-10-369-493-15690
 ; Sequence 15690, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 15690
 ; LENGTH: 276
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas campestris
 ; US-10-369-493-15690

Query Match 38.4%; Score 582; DB 12; Length 276;
 Best Local Similarity 46.0%; Pred. No. 1.4e-54;
 Matches 126; Conservative 41; Mismatches 105; Indels 2; Gaps 2;

QY 11 WPAPAKLMLRTGRRPDGTHLLQTVFQMLDCLMTFHPVDDGRV-TLRNPISGVPEQ 69
 Db 3 WSPAPAKLMLRTGRRPDGTHLLQTVFQMLDCLMTFHPVDDGRV-TLRNPISGVPEQ 62
 QY 70 DLTVRAANLTKSHGCVGVCIDIEKNLPMGGGLGGSSDAATTVLNLMLG 129
 Db 63 DDLWRAARALQIHAGTALGAEIRYDKRIPAGGFGGSSDAATTVLNLMLG 122
 QY 130 ELMDELRLGADVPEVFGCSAMGSEVEDLQATLPEQWFIKKPDCHVNTGE 189
 Db 123 TLAEGLRLGADVPEVFGKHAFAEGIGEKLRPASPEEKPLVHPGISISTPL 182
 QY 190 LTRNSAAVTMSDFLAGNNDCESEVCKLYRPVDAI-DALLCYAERLTGTGAC 248
 Db 183 LTRDAAPAKIADPAGSSLLDNAPFVLRREPAIEAVFQALSRIGTRLTGSGSCFVEF 242
 QY 249 CCKEDAESALBGLKDRWLVFLAKLNOSALY 282
 Db 243 ATRAAAEQAMHLPENLRAMWVEGAHSPLLDALD 276

RESULT 10
 US-10-369-493-16082
 ; Sequence 16082, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 16082
 ; LENGTH: 276
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas campestris
 ; US-10-369-493-16082

Query Match 38.4%; Score 582; DB 12; Length 276;

Best Local Similarity 46.0%; Pred. No. 1.4e-54;
Matches 126; Conservative 41; Mismatches 105; Indels 2; Gaps 2

Qy	11	WPAKAKINMLRRTGRRPGCYHLQVPPOMLQCMWLEFHPVDSRV-TLKNPIISGVEQ	69
Db	3	WPAKALNLELQIVGRBBDYTHLQVFEFLDPMSTYHAYRLTDQQLRIGASLPGVAED	62
Qy	70	DDLTVRAANLKSHTGCVRGVCIDIEKMLPMGGSLGGSSDAATTLVVLNLMGLGSKR	129
Db	63	DDLTVRAAALQTHAFSTALGAELRDYKRIIPAGGGGCGSSDAATVLYALNLMGLGFPVD	122
Qy	130	ELMDLGRIGADVPVPPGCSAMGSGVSDDLAITLPPOMVIIKPDCHVATGELPSAEN	189
Db	123	TLMEELGRIGADVPVPPGHNMAAEVGEKLLPISLPDAAYLVDPGIIHPTPVLPSQSE	182
Qy	190	LITNSAAVTVSDFLIAGDNNDCSEVVCIKLYREVKDAI-DALLCYAEALITGTGA CVFAQF	248
Db	183	LITDAAAPAKIAVPASGSLDINAFFEVLRRREPAIEAVPQALSRIETPLTSGSGCEVEF	242
Qy	249	CNKEDAESALBGLKDWILFLAKGINSQSLATYKLL	282
Db	243	ATPDAAAEQMAHLPGILRAVVEGAAHSLIEDAL	276

```

RESULT 11
US-10-369-493-17751
Sequence 17751, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Jiongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 17751
LENGTH: 298
TYPE: PRT
ORGANISM: Xylella fastidiosa
US-10-369-493-17751

```

Query Match	38.3%	Score 580.5;	DB 12;	Length 298;
Best Local Similarity	46.3%	Pred. No. 2.3e-54;		
Matches 132;	Conservative 44;	Mismatches 106;	Indels 3;	Gaps 3

Qy	1	MDVAAAGSEKMPAPAKINLMLRTTGRPRGYHLLQVFPOMLDCMLFPHVDGRV-TL	59
Db	11	VDSQVGH-SAMPAKINLNFLOTITGRVRYGHLEQLVFPLLMGTHILRRBEDQIHRI	69
Qy	60	RNPISVPEODDITVPAANILKSHGSCVAGVICIDIEKNLPMGSGGGSSDAATLVVTLN	113
Db	70	GESVTVGVAEADDLVVPAAYLILKATVWHLGAIDFVEKRI.PVGGGFGGSSDAATVLVTLN	122
Qy	120	RLMGLGSKRELMDIGRLGADVPVPPGCSAMGSEVSEDLQALITLPQWFIILKPDCHV	179
Db	130	ALMHTRLDVAVLNALSRLGADVPVPHGCSAMABSGVGLTPEMLLPQAATLIDPGVCV	188
Qy	180	NIGEIFSAENLTENSAVVTMSDFLAGDNNDGSE-VVCHLYRVPDAIDALLCYAEARLT	238
Db	190	PTRELFLPDLPTDPAAPATIGDIAGTAFAGNAFEPVLRARESAVAGALDVLSEVGPAVTE	244
Qy	239	GTGCAFCACFCNKEDAESALEGKQDNVLVFLAKLNGSLNYQDCE	283
Db	250	GSGGGCGVEFTSDDEACALERTIPYGLCAMVADGASRSFLDLVTK	294

RESULT 12
US-10-369-493-9184
; Sequence 9184, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:

```

1  APPLICANT: Gao, Yongwei
2  APPLICANT: Hinkle, Gregory J.
3  APPLICANT: Slater, Steven C.
4  APPLICANT: Goldman, Barry S.
5  APPLICANT: Chen, Xianfeng
6  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
7  TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
8  FILE REFERENCE: 38-10(52052)B
9  CURRENT APPLICATION NUMBER: US/10/369,493
10 CURRENT FILING DATE: 2003-02-28
11 PRIOR APPLICATION NUMBER: US 60/360,039
12 PRIOR FILING DATE: 2002-02-21
13 NUMBER OF SEQ ID NOS: 47374
14 SEQ ID NO 9184
15 LENGTH: 270
16 TYPE: PRT
17 ORGANISM: Xylella fastidiosa
18 US-10-369-493-9184

```

Query Match	37.7%;	Score 572;	DB 12;	Length 270;
Best Local Similarity	47.4%;	Pred. No. 1.7e-53;		
Matches 128;	Conservative 42;	Mismatches 98;	Indels 2;	Gaps 2

QY 1 WPAPALNTAMLTITGRPEPGYTLQTVQM.DLCWLTFFHVDGRV-TLNNPISGVEQ 69
 Db 1 WPAPALNT.FLOTITRRRDYGHETQTVRLDMGDTTHLRFBEGQ.IHRIGESVTVVEA 60
 QY 70 DDI.TVRANLTAKSHGCVGVCIDIEKQLPMGGIGGSSDAATTVLVANRLMGLGSKR 123
 Db 61 DDI.VVRAAAILTHAINTVYIGADIPFEKRIPIVGGGFGGSSDAATVLLVLANLMTETRDVA 120
 QY 130 EIMDGLRTAGADVPYVFGSGAMGGSVEEDIQATLTPQWVITIKPOCHVNTGIEFSAEN 189
 Db 121 VILNALGRLRGADVPYVFGGCAAMAGSVECLTPMLIRGAAYLLDPCVCPTRELFDPD 180
 QY 190 LTRNSAVYTMSDFLAGDNRNDQSE-VYCKLYRPYKDAIDALLCYAEARLTGTGACVPAQF 248
 Db 181 LTRDASPAITGFIAGAAFGNAFPEVTLRRRESAAVAGALDVLBSYGPAVTVGSGGCFVEF 240
 QY 249 CNKEDASALSGTLKDRMTLVFLAKINDSAL 278
 Db 241 STRDASCKLERTLPYGLCAWADGASRSEPL 270

```

RESULT 13
US-10-369-493-9359
; Sequence 9359, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9359
; LENGTH: 270
;
; TYPE: prt
; ORGANISM: Xylella fastidiosa
; IS-10-369-493-9359

```

Query Match 37.7%; Score 572; DB 12; Length 270;
 Best Local Similarity 47.4%; Pred. No. 1.7e-53;
 Matches 128; Conservative 42; Mismatches 98; Indels 2; Gaps 2;

QY 11 WPAPAKNIMLRITGRPRDGYHLLQTVFQMDLCMLTFHPVDDGRV-TLRNPISGVEEQ 69
 DB 1 WPAPAKNIMLRITGRPRDGYHLLQTVFQMDLCMLTFHPVDDGRV-TLRNPISGVEEQ 60
 QY 70 DDLTVRANMLKSHGCVRGVCIDIBKXNPMWGGGSGSSDAATLVVLRNLMGJLSKR 129
 DB 61 DDLTVRANMLKSHGCVRGVCIDIBKXNPMWGGGSGSSDAATLVVLRNLMGJLSKR 120
 QY 130 BLMDGLRLGADVPPVFGCSAMGEGVEDLQATITLPEQWFIIPKDCHVNTGELIFSAN 189
 DB 121 VLAALGLRLGADVPPVFGCSAMGEGVEDLQATITLPEQWFIIPKDCHVNTGELIFSAN 180
 QY 190 LTRNSAVVTMSDFLAGDNNDCE-VYCKLTPRYKDAIDALICVBAALITGTACVPAQF 248
 DB 181 LTRNSAVVTMSDFLAGDNNDCE-VYCKLTPRYKDAIDALICVBAALITGTACVPAQF 240
 QY 249 CNKEDAESALBGLKXDRWLVFLAKGNOSAL 278
 DB 241 STRDEKCHLRLPYGLCAWVADGASRSEPL 270

RESULT 14 US-10-369-493-19675

; Sequence 19675, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 19675
 ; LENGTH: 213
 ; TYPE: PRT
 ; ORGANISM: Nitrosomonas europaea
 US-10-369-493-19675

Query Match 36.7%; Score 556; DB 12; Length 213;
 Best Local Similarity 51.6%; Pred. No. 6.6e-52;
 Matches 110; Conservative 34; Mismatches 67; Indels 2; Gaps 2;

QY 11 WPAPAKNIMLRITGRPRDGYHLLQTVFQMDLCMLTFHPVDDGRVTLRNPISGVEEQ 70
 DB 1 WPAPAKNIMLRITGRPRDGYHLLQTVFQMDLCMLTFHPVDDGRVTLRNPISGVEEQ 60
 QY 71 DDLTVRANMLKSHGCVRGVCIDIBKXNPMWGGGSGSSDAATLVVLRNLMGJLSKR 129
 DB 61 DDLTVRANMLKSHGCVRGVCIDIBKXNPMWGGGSGSSDAATLVVLRNLMGJLSKR 120
 QY 130 BLMDGLRLGADVPPVFGCSAMGEGVEDLQATITLPEQWFIIPKDCHVNTGELIFSAN 189
 DB 121 VLAALGLRLGADVPPVFGCSAMGEGVEDLQATITLPEQWFIIPKDCHVNTGELIFSAN 180
 QY 190 LTRNSAVVTMSDFLAGDNNDCE-VYCKLTPRYKDAIDALICVBAALITGTACVPAQF 248
 DB 181 LTRNSAVVTMSDFLAGDNNDCE-VYCKLTPRYKDAIDALICVBAALITGTACVPAQF 240

RESULT 15
 US-10-369-493-8419
 ; Sequence 8419, Application US/10369493

; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 8419
 ; LENGTH: 233
 ; TYPE: PRT
 ; ORGANISM: Ralstonia metallidurans
 US-10-369-493-8419

Query Match 35.2%; Score 533; DB 12; Length 233;
 Best Local Similarity 49.1%; Pred. No. 2.4e-49;
 Matches 114; Conservative 35; Mismatches 77; Indels 6; Gaps 2;

QY 53 DDGRVTLRNPISGVEEQDULTVRANMLKSHGCVRGVCIDIBKXNPMWGGGSGSSDA 112
 DB 2 DDGRVTLRNPISGVEEQDULTVRANMLKSHGCVRGVCIDIBKXNPMWGGGSGSSDA 61
 QY 113 TTVVLRNLMGJLSKRDLGLRLGADVPPVFGCSAMGEGVEDLQATITLPEQWFI 172
 DB 62 TTVVLRNLMGJLSKRDLGLRLGADVPPVFGCSAMGEGVEDLQATITLPEQWFI 121
 QY 173 IKPCHVNTGELIFSANLITNSAVVTMSDFLAGDN-----RNDCEVYCKLTPRYKDAID 227
 DB 122 IKPCHVNTGELIFSANLITNSAVVTMSDFLAGDN-----RNDCEVYCKLTPRYKDAID 181
 QY 228 ALICVBAALITGTACVPAQFQCNKEDAESALBGLKXDRWLVFLAKGNOSAL 278
 DB 182 ALICVBAALITGTACVPAQFQCNKEDAESALBGLKXDRWLVFLAKGNOSAL 233

Search completed: January 29, 2004, 16:21:17
 Job time : 20.4547 secs

DR N-PSDB; ABRK3266.

XX New high growth methanotrophic bacterial strain, useful for producing
PT single cell proteins, grows on a C1 carbon substrate, and comprises a
PT functional gene encoding in Embden-Meyerhof carbon pathway

XX Claim 11; Page 149-150; 157pp; English.

XX The invention relates to a high growth methanotrophic bacterial strain,
CC which grows on a C1 carbon substrate e.g. methane and methanol, and
CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a
CC gene coding a pyrophosphate dependent phosphotransferase enzyme or a 16S
CC rRNA. The bacterial strain is useful for the production of single cell
CC protein and for the biotransformation of a nitrogen-containing compound,
CC e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the
CC production of a feed product comprising a protein, carbohydrates and a
CC pigment and for reducing oxygen demand, for removing nitrates and
CC nitrites in methane-containing environments such as landfills, waste
CC water treatment systems or anywhere that methane, oxygen and nitrates are
CC present. The bacterial strain of the invention can be used as a
CC denitrifying agent for the conversion of nitrate or nitrite to nitrous
CC oxide with methane or methanol as a carbon source. It is also used in the
CC production of biomass including proteins, carbohydrates and a wide
CC variety of pigments (particularly for isoprenoid pigments for the
CC purpose of generating animal feeds), in production of terpenoid and
CC carotenoid compounds, useful as pigments and as monomers in polymeric
CC materials and in production of exopolysaccharides at high levels.
CC Sequences ABG61551-ABG61590 represent high growth methanotrophic
CC bacterial strain proteins of the invention.

XX Sequence 285 AA;

Query Match 100.0%; Score 1516; DB 23; Length 285;

Best Local Similarity 100.0%; Pred. No. 1.3e-155;

Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MDYAAAGGGERPAPAKLMLRITGRPDGYHLQTVFQMLDLCDFWTFHPVDGRTLR 60
DB 1 MDYAAAGGGERPAPAKLMLRITGRPDGYHLQTVFQMLDLCDFWTFHPVDGRTLR 60
QY 61 NPISGVPBQDDLTVRANLILKSHTCVRCVCIIDIEKNLPMGGGLGGSSSDAATTLVLNR 120
DB 61 NPISGVPBQDDLTVRANLILKSHTCVRCVCIIDIEKNLPMGGGLGGSSSDAATTLVLNR 120
QY 121 LMGLGSKRELMDLGRIGADVPVFVFGCSAMGEGVEDLQATILPEQWFTIIRPDCHVN 180
DB 121 LMGLGSKRELMDLGRIGADVPVFVFGCSAMGEGVEDLQATILPEQWFTIIRPDCHVN 180
QY 181 TGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVCAIDALCYAEARLTGT 240
DB 181 TGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVCAIDALCYAEARLTGT 240
QY 241 GACVFAQFCNKEDASALEGLKDRMLVFLAKGLNQSALYKYLEOG 285
DB 241 GACVFAQFCNKEDASALEGLKDRMLVFLAKGLNQSALYKYLEOG 285

```

RESULT 2

AAE22304

ID AAE22304 standard; Protein; 285 AA.

XX AAE22304;

XX 25-JUL-2002 (first entry)

XX Methylomonas 16a sp. 4-diphosphocytidyl-2-C-methylerythritol kinase.

XX Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;
KM anti-oxidant; steroid; flavour; fragrance; electro-optic application;
KM aquaculture; enzyme; 4-diphosphocytidyl-2-C-methylerythritol kinase;
ISGB.

OS Methylomonas 16a sp.

XX WO200218617-A2.

XX 07-MAR-2002.

XX 04-SEP-2001; 2001WO-US27420.

XX 01-SEP-2000; 2000US-229858P.

XX 01-SEP-2000; 2000US-229807P.

XX (DUPLO) DU PONT DE NEMOURS & CO E. I.

XX Brostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX Odum JM, Picaltaggio SK, Rouviere PE;

XX WPI; 2002-351711/38.

XX N-PSDB; AAD35502.

XX Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by
PT using microorganisms having a nucleic acid molecule encoding enzymes in
PT the carotenoid biosynthetic pathway and which metabolize single carbon
PT substrates

XX Claim 44; Page 116-117; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds.
CC The method comprises a transformed metabolising host cell, comprising
CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule
CC encoding an enzyme in the carotenoid biosynthetic pathway, under the
CC control of regulatory sequences; and contacting the host cell with carbon
CC substrate to produce a carotenoid compound. The method is useful for
CC producing carotenoid compounds such as antheraxanthin and astaxanthin, by
CC using microorganisms having a nucleic acid molecule encoding enzymes in
CC the carotenoid biosynthetic pathway and which metabolize single carbon
CC substrates. The carotenoids have potent anti-oxidant properties useful in
CC diet, and aquaculture elements. The carotenoids are also useful as
CC intermediates in the synthesis of steroids flavours and fragrances and
CC compounds for potential electro-optic applications. The present sequence
CC is Methylomonas 16a sp. 4-diphosphocytidyl-2-C-methylerythritol kinase
CC (isgb) enzyme used in the invention.

XX Sequence 285 AA;

Query Match 100.0%; Score 1516; DB 23; Length 285;

Best Local Similarity 100.0%; Pred. No. 1.3e-155;

Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MDYAAAGGGERPAPAKLMLRITGRPDGYHLQTVFQMLDLCDFWTFHPVDGRTLR 60
DB 1 MDYAAAGGGERPAPAKLMLRITGRPDGYHLQTVFQMLDLCDFWTFHPVDGRTLR 60
QY 61 NPISGVPBQDDLTVRANLILKSHTCVRCVCIIDIEKNLPMGGGLGGSSSDAATTLVLNR 120
DB 61 NPISGVPBQDDLTVRANLILKSHTCVRCVCIIDIEKNLPMGGGLGGSSSDAATTLVLNR 120
QY 121 LMGLGSKRELMDLGRIGADVPVFVFGCSAMGEGVEDLQATILPEQWFTIIRPDCHVN 180
DB 121 LMGLGSKRELMDLGRIGADVPVFVFGCSAMGEGVEDLQATILPEQWFTIIRPDCHVN 180
QY 181 TGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVCAIDALCYAEARLTGT 240
DB 181 TGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVCAIDALCYAEARLTGT 240
QY 241 GACVFAQFCNKEDASALEGLKDRMLVFLAKGLNQSALYKYLEOG 285
DB 241 GACVFAQFCNKEDASALEGLKDRMLVFLAKGLNQSALYKYLEOG 285

```

RESULT 3

AAU80328

ID AAU80328 standard; Protein; 285 AA.

XX AAU80328;

XX		15-JUL-2002	(first entry)
DT			
XX		Methylomomas 16a ORF4 ychB/ispe protein sequence.	
DB			
XX		Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;	
KM		keratenoid; pigment; flavour; fragrance; open reading frame 4; ORF4;	
KN		yehB; 4-diphosphocytidylyl-2-C-methylethylthiol kinase enzyme; ispe.	
XX			
OS		Methylomomas sp.	
XX			
PN		WO200220733-A2.	
FD			
XX		14-MAR-2002.	
PF			
XX		29-AUG-2001; 2001WO-US26852.	
PR			
XX		01-SEP-2000; 2000US-229907P.	
PA			
XX		(DUPO) DU PONT DE NEMOURS & CO E I.	
PI			
P1		Cheng Q, Koffas M, Norton KC, Odom JM, Picataggio SK, Rouviere PB;	
XX		Schenzle A, Tomb J,	
DR		WP1; 2002-383051/41.	
XX		N-PSDB; ABK50084.	
PT			
PT		Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,	
PT		isolated from Methylomomas 16a, useful for the production of isoprenoid	
XX		compounds -	
PS			
XX		Claim 4; Page 71-72; 84pp; English.	
CC		The present invention relates to a new nucleic acid molecule encoding	
CC		an isoprenoid biosynthetic enzyme isolated from Methylomomas 16a.	
CC		The invention is useful for obtaining a nucleic acid molecule	
CC		encoding an isoprenoid compound biosynthetic enzyme, and for the	
CC		microbial production of isoprenoid compounds. The molecules of the	
CC		invention are also useful for regulating isoprenoid biosynthesis in an	
CC		organism and for producing recombinant organisms for producing various	
CC		isoprenoid compounds. The nucleic acid is also useful for feed additive,	
CC		for the production of keratenoids and their derivatives, isoprenoid	
CC		intermediates, and as pure products useful as pigments, flavours and	
CC		fragrances. The present amino acid sequence represents the Methylomomas	
CC		16a open reading frame 4 (ORF4) ychB/ispe (4-diphosphocytidylyl-2-C-	
CC		methyleneithiol kinase enzyme) protein of the invention, as described	
CC		above.	
SC			
Sequence	285 AA;		
Query Match:	100.0%; Score 1516; DB 23; Length 285;		
Best Local Similarity	100.0%; Pred. No. 1.3e-155;		
Matches	285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 MDVAAAGMEERWAPAKLNMAMRITRRPDGYHLLQTVOMQLDLCMLTFHPVDGRVTLR	60	
DB	1 MDVAAGMEERWAPAKLNMAMRITRRPDGYHLLQTVOMQLDLCMLTFHPVDGRVTLR	60	
QY	61 NPISGVPEDDLLTVRAANLLKSHGTGCVRGVCIDIRKNLPWGSGTGGSSDAATTLVLIAR	120	
DB	61 NPISGVPEODDLTVRAANLLKSHGTGCVRGVCIDIRKNLPWGSGTGGSSDAATTLVLIAR	120	
QY	121 LMGGLSLGRELMDDGLRLGADVPPVFPGCSAAGBSVSDDLQATLLPEGMFYIIRPDCHVN	180	
DB	121 LMGLGLSLGRELMDDGLRLGADVPPVFPGCSAAGBSVSDDLQATLLPEGMFYIIRPDCHVN	180	
QY	181 TGEIFSAENLTRNSAVVTMSDFLAGDNNDNSEVVCXKLRYPKADIALALCYAEARLTGT	240	
DB	181 TGEIFSAENLTRNSAVVTMSDFLAGDNNDNSEVVCXKLRYPKADIALALCYAEARLTGT	240	
QY	241 GACVFAOFCNKEDAESALEGKKORPLVLAAGLNLSALYKYLEEG	285	
DB	241 GACVFAOFCNKEDAESALJGLDKRPLVLAKGLNLSALYKYLEEG	285	

ID	ABP78450	standard; Protein; 281 AA.
AC	ABP78450;	
DT	07-MAR-2003	(first entry)
DE	N. gonorrhoeae amino acid sequence SEQ ID 3430.	
XX	Antibacterial; infection; vaccine; gene therapy.	
XX	Neisseria gonorrhoeae.	
XX	WO200279243-A2.	
XX	10-OCT-2002.	
XX	13-FEB-2002; 2002WO-IBO2069.	
XX	12-FEB-2001; 2001GB-0003424.	
XX	(CHIR-) CHIRON SPA.	
XX	Fontana MR, Pizza M, Maignani V, Monaci E,	
XX	KPI; 2003-058415/05.	
XX	N-PSDB; ABZ39420.	
XX	New protein from Neisseria gonorrhoeae, useful for the manufacture of a	
XX	medicament for treating or preventing N. gonorrhoeae infection	
XX	Disclosure; Page 439; 815pp; English.	
XX	The present invention relates to proteins from Neisseria gonorrhoeae.	
XX	Also disclosed are the nucleic acid molecules encoding the proteins and	
XX	antibodies that specifically bind to the proteins. The composition	
XX	comprising the protein, nucleic acid or antibody is useful for the	
XX	manufacture of a medicament for treating or preventing N. gonorrhoeae	
XX	infection, this may be in the form of a vaccine or gene therapy.	
XX	Sequences given in records ABP76736-ABP81046 represent nucleic acid	
XX	molecules of the invention.	
XX	Sequence 281 AA;	
XX	Query Match 41.2%; Score 624; DB 24; Length 281;	
XX	Best Local Similarity 46.1%; Pred. No. 7, 1e-59;	
XX	Matches 130; Conservative 54; Mismatches 90; Indels 8; Gaps 3;	
QY	1 MDVAAGSGERPAKINLMKRTGRRPDGYHLLQTVFQMLDCDMLTFEYVDGRVTLR 60	
DB	1 MNINDG-RQAFPAFAKINLDIRLTGRREDGHNIESIFCLIDLDYTLKPRDGIILH 59	
QY	61 NPISGVEPDLDITRAANILKSHHGCVRGVCIIDIEKTLPMGGGLGGSSDPATLVTLNR 120	
DB	60 NPVGRIQSEADLSIRASLSLQKTRAKNLAGEVIMIDKIFPGAGLGGSSDPATLVTLNLR 119	
QY	121 LMEGLSKREIMDLGRLGADVPEVFGCSAMGSEVSEDIQATLTPQDMFVIIRKPCVNR 180	
DB	120 WMQGLTQWGLDIDGALGADVPEFPIFGKNAFASGIGKILIGMDIPKQMYIVKSPVHVS 179	
QY	181 TGEIFSAENLIRNSAVVTMSDFLAGDN----RNDCSVWCKLRPVMDAIDALICVAEAR 236	
DB	180 TAKITFTEGLTRDASSIMPTF--QNLQPPRNDQAVVREKIPSVKAYSELKYSQAM 236	
QY	237 LVTGACVCPAQCCKEDAESALBGLKQRMVLVFLAKIGNQAL 278	
DB	237 MTSGGACIFAAFOARNSAVNIYIVQVSLYDAAYLABGISKAPL 278	

AAB11374
ID AAB11374 standard; protein; 283 AA.

AC AAB11374;

XX 22-FEB-2001 (first entry)

XX B. coli YCHB protein.

DE YJEB; KDTB; YOGF; YGGB; YHBC; YGBB; YCHB; antibacterial;
treatment; infection.

XX Escherichia coli.

XX DE19916176-A1.

XX 12-OCT-2000.

XX 10-APR-1999; 99DE-1016176.

XX 10-APR-1999; 99DE-1016176.

XX (FARB) BAYER AG.

XX Broetz H, Ehler K, Freidberg C, Spallmann F, Wieland B;

XX Labischinski H;

XX WPI; 2000-639611/62.

XX N-PSDB; AAC66044.

XX Essential genes from bacteria, useful in screening for antimicrobial
agents, and related proteins, transformants and antisense sequences -

XX Disclosure; Page 23-24; 28pp; German.

CC This invention describes novel Escherichia coli genes (i) encoding
proteins (iii) designated YOGF, YHBC, YGGB, YCHB, YGBB, YJEB and
KDTB, and genes (ia) that encode orthologous gene products (iia) in
other microorganisms and which have antibacterial activity. Recombinant
microorganisms in which expression of (i) or (ia) can be regulated are
used to identify compounds that bind to the gene products, particularly
in affinity selection assays. (ii) and (iia) are used to identify, or
prepare, antibodies and other proteins that bind to the gene products.
CC Substances that bind to (ii) or (iia) are potentially useful as
antibacterials for treating a wide range of infections in humans and
CC animals. Sequences antisense to (i) and (ia) can also be used as
antibacterials. The specified genes are widely distributed in bacteria
CC but have no close homologs in eukaryotic cells.

XX Sequence 283 AA;

Query Match 40.6%; Score 616; DB 21; Length 283;
Best Local Similarity 47.9%; Pred. No. 5.3e-58;

Matches 135; Conservative 39; Mismatches 96; Indels 12; Gaps 4;

10 RMPAPAKLNMRLITGRPRDGYHLQTVFQMLDLCDFTHFPVDDGRVTLNPIISGVPEQ 69

4 QMPSPAKNLFYITGRADGYHLQTVFQMLDLCDFTHFPVDDGRVTLNPIISGVPEQ 63

70 DDLTVRAANL-----KSHTCVRCVCIIDENKLMFGGGLGGSSDAATTLVNLRL 121

64 DNLIVRAARLTKTAADSGRLPTG--SGANISIDKRLPMGGGLGGSSNAATLVNLRL 121

122 MGIGLSKRELDLGLRIGADVVPVFGCSAMGEGVSEDLQATILPEQFVILIKDDCHVNT 181

122 WQCGLSMDLAEWGLTIGADVVPVFGGHAAPAEVGEILITPVDPEKXVLVAHPGVS IPT 181

182 GEIFSSENTLINSVAVTMSDFLAGDNRCSEVVCGLYRPVKDAIDALLCYAERLTGTG 241

182 PVIRPDPLPRLPNTPKRSIETLLKCFSPNDCEVIARKRREVDVATLMLLEVAERLTGTG 241

242 ACVPAQPFNKEDASALEGLKDRNL-VFLAKGLNQSALYKTL 282

DB 242 ACVPAQPFNKEDASALEGLKDRNL-VFLAKGLNQSALYKTL 282

RESULT 6

ID AAB68286 standard; Protein; 283 AA.

XX AAB68286;

XX 09-JUL-2001 (first entry)

XX Amino acid sequence of an isopentenyl monophosphate kinase.

XX Isopentenyl monophosphate kinase; IPK; isoprenoid biosynthesis; pigment;

XX vitamin; essential oil; pathogen resistance; antibiotic; herbicide;

XX antimicrobial.

XX Escherichia coli.

XX W0200132907-A1.

XX 10-MAY-2001.

XX 02-NOV-2000; 2000MO-US30289.

XX 04-NOV-1999; 99US-0434774.

XX (UNIV) UNIV WASHINGTON STATE RES FOUND.

XX Croteau RB, Lange BM;

XX WPI; 2001-308747/32.

XX N-PSDB; AAF85079.

XX New nucleic acid encoding isopentenyl monophosphate kinase, useful e.g.
for making transgenic plants with increased synthesis of isoprenoids,

XX e.g. essential oils -

XX Example 1; Page 48-49; 62pp; English.

CC The present sequence represents an isopentenyl monophosphate kinase (IPK)
polypeptide. The enzyme is an isoprenoid biosynthesis stimulator. The
IPK polynucleotide is useful recombinant production of IPK, as a source
of probes, primers and antisense sequences and for increasing/reducing
CC expression levels of IPK in cells, particularly of essential oil
CC plants, so as to increase flow through the isoprenoid biosynthesis
CC pathway, resulting in increased production of e.g. pigments, vitamins
CC and essential oils, also to increase resistance to pests and pathogens
CC and to improve plant strength. Mutant forms of the IPK polynucleotide
CC can be used to express forms of IPK that are resistant to IPK-targeted
CC herbicides, and recombinant IPK can be used to screen for antibiotics,
CC herbicides and antimicrobial agents directed against IPK.

XX Sequence 283 AA;

Query Match 40.6%; Score 616; DB 22; Length 283;
Best Local Similarity 47.9%; Pred. No. 5.3e-58;

Matches 135; Conservative 39; Mismatches 96; Indels 12; Gaps 4;

10 RMPAPAKLNMRLITGRPRDGYHLQTVFQMLDLCDFTHFPVDDGRVTLNPIISGVPEQ 69

4 QMPSPAKNLFYITGRADGYHLQTVFQMLDLCDFTHFPVDDGRVTLNPIISGVPEQ 63

70 DDLTVRAANL-----KSHTCVRCVCIIDENKLMFGGGLGGSSDAATTLVNLRL 121

64 DNLIVRAARLTKTAADSGRLPTG--SGANISIDKRLPMGGGLGGSSNAATLVNLRL 121

122 MGIGLSKRELDLGLRIGADVVPVFGCSAMGEGVSEDLQATILPEQFVILIKDDCHVNT 181

122 WQCGLSMDLAEWGLTIGADVVPVFGGHAAPAEVGEILITPVDPEKXVLVAHPGVS IPT 181

182 GEIFSSENTLINSVAVTMSDFLAGDNRCSEVVCGLYRPVKDAIDALLCYAERLTGTG 241

DB 182 PVIFKDPKELPNTPKRSIETLLKCEFSNDCEVIARFREVDAVLSMLEYAPSLNCTG 241

QY 242 ACVFAQFCNKEDASALBGLDRML-VFLAKGLNSALYKEL 282

DB 242 ACVFAEPDTESEARQVLEQAPL-WLNGVFAKGANISPLHRAM 282

RESULT 7

ID ABP25586 standard; Protein; 283 AA.

XX ABP25586;

AC ABP25586;

DT 02-JUL-2002 (first entry)

XX

DE Streptococcus polypeptide SEQ ID NO 348.

XX

KM Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

XX

PN KO200234771-A2.

XX

PD 02-MAY-2002.

XX

PF 29-OCT-2001; 2001WO-GB04789.

XX

PR 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

XX

PA (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

PI Telford J, Masigiani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tettelin H;

XX

DR WPI; 2002-352536/38.

DR N-PSDB; ABN66217.

XX

PT New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

XX for detecting a compound that binds to the protein.

PS Claim 1; Page 3188; 452536p; English.

XX

CC The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

XX

XX Sequence 283 AA;

XX

Query Match 21.9%; Score 332; DB 23; Length 283;

Best Local Similarity 32.2%; Pred. No. 3,4e-27;

Matches 85; Conservative 58; Mismatches 107; Indels 14; Gaps 5;

QY 13 APAKINLMIRITGRBDDGYHLLQTVFQMLDLCMLTFHPVDGVTLLNPIISGVPEODDL 72

DB 7 APAKINLMIRITGRBDDGYHLLQTVFQMLDLCMLTFHPVDGVTLLNPIISGVPEODDL 66

QY 73 TV-PAANLLKSHNGCVAGVCIDIEKNLPMGGGLGGSSDAATTLVNLRLMGLSREL 131

DB 67 DVEFGADIIKQGYINNGVHRIEKSIPVCAIGGGSTDAATRLALNRLMNLQMDYDEM 126

QY 132 MDGLREAGADVPRVPRCCSAMGCVSDDQAL-TLPEQWVITIKPDCHVNTGEF---SA 187

DB 127 VAIGFKISDVPICLGCCSLVJGKGIYKPLPTLRPCWLVLYVDPDGISTKSTFRIDIC 186

QY 188 ENLIR-----NSAVVMSDFLAGDNEMDCSEVCKLYREKDAIDALLCYA---EARLT 238

DB 187 KSISRVDIDLKSAIISDQYQVAMKSMGNSLEDITTKRNVISITIKERMLNSGADVALMT 246

QY 239 GTGACVPAQFCNKEDASALBGLK 262

DB 247 GSGPTVFSMCKTEKKADRVNSMK 270

RESULT 8

ID ABB50111 standard; Protein; 293 AA.

XX ABB50111;

AC ABB50111;

XX

DT 05-FEB-2002 (first entry)

XX

DE Listeria monocytogenes protein #2815.

XX

KM Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

KW vitamin B12; bacterial infection; disease.

XX

OS Listeria monocytogenes.

XX

PN KO200177335-A2.

XX

PD 13-OCT-2001.

XX

PF 11-APR-2001; 2001WO-FR01118.

XX

PR 11-APR-2000; 2000FR-0004629.

XX

PA (INSP) INST PASTEUR.

XX

PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Feihl H, Dehoux P;

PI Dussauguet O, Cheuouani F, Nedjari H, Glaser P, Kunst F, Cossart P;

PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;

PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;

PI Chakraborty T, Doanann E, Hain T, Berche P, Chardot A, Durant L;

PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

PI Matuteo E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;

PI Rose M, Voss H;

XX

DR WPI; 2002-010914/01.

XX

PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment

PT and prevention of Listeria and related bacterial infections, and

XX related polypeptides.

PS Claim 6; SEQ ID NO 2816; 192pp; French.

XX

CC The present invention relates to the genome sequence of Listeria

CC monocytogenes BDD-e (see ABA03041). The genome sequence and fragments of

CC it are useful for selecting probes and primers for detecting genes in L.

CC monocytogenes and related organisms, and for studying genetic

CC polymorphisms and other genomes. The present sequence is a protein

CC encoded by the genome sequence of the present invention. Proteins

CC expressed from the genome sequence are useful for raising specific

CC antibodies, identification of L. monocytogenes and related organisms, and

CC for biosynthesis and biodegradation, especially biosynthesis of vitamin

CC B12. The genome sequence and proteins encoded by it are also useful for

CC selecting compounds that regulate gene expression and cell replication

CC and modulate *L. monocytogenes*-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by *L.*
 CC *monocytogenes* and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pcc_sequences.

CC Sequence 293 AA;

Query Match 21.1%; Score 320; DB 23; Length 293;

Best Local Similarity 33.7%; Pred. No. 7, 1e-26;
 Matches 89; Conservative 41; Mismatches 120; Indels 14; Gaps 6;

```

QY 13 ABAKINMLRITGRPRDGYHLQTVFQMLDLCMTFFHPVDGRVTLNPNISGVE-ODD 71
DB 7 ABAKINMLSDALYKXEDGHEVEVMWTTIDLADRLYLBRDEBKLYLVKXHPFPERERN 66
QY 72 LTVRAANILKSHTGCVRCVCIIDIEKQLPMGGAGGGSSDAATTLVNLRLMGLSKREL 131
DB 67 LTVRAANILKSHTGCVRCVCIIDIEKQLPMGGAGGGSSDAATTLVNLRLMGLSKREL 126
QY 132 MDLGRLAGADVPVPGCSAMGEGVEDLOAI-TLPGQVYIIKPDCHVNTGEIF--SA 187
DB 127 AEISSEIGSDIAFCYGGTALATGEGKISALPNIPGCVIYANGSISVSTIYKEIQV 186
QY 188 ENLTNSAVVWMSDFLAGD-----NNDCEVVCCKLYRPYKCAIDALLCY-ABARL-7 238
DB 187 DNEVEHPDQKMIIESKNGDLDGIFATGNTVLESVLEKPNQVKIKDMLAFGAALMS 246
QY 239 GTGACVFAQFCNKEDEASALEGLK 262
DB 247 GSGPTVFALIKQYSPKRVNGLR 270

```

RESULT 9

ABP38503
 ID ABP38503 standard; Protein; 299 AA.

AC ABP38503;

DT 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3348.

KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KM antibacterial; gene therapy.

OS Staphylococcus epidermidis.

PN US6380370-B1.

PD 30-APR-2002.

PF 13-AUG-1998; 98US-0134001.

PR 14-AUG-1997; 97US-055779P.

PR 08-NOV-1997; 97US-064964P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;

DR MPI; 2002-381255/41.

DR N-PSDB; ABN91048.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 PS Disclosure; SEQ ID 3348; 267bp; English.

CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences

CC given in ABP3124 to ABP37960. The *S. epidermidis* sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly *S. epidermidis* infections. The sequences can be used to
 CC screen for compounds able to interfere with the *S. epidermidis* life
 CC cycle or inhibit *S. epidermidis* infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.

CC Sequence 299 AA;

Query Match 19.9%; Score 302; DB 23; Length 299;
 Best Local Similarity 32.0%; Pred. No. 6, 5e-24;
 Matches 79; Conservative 51; Mismatches 103; Indels 14; Gaps 6;

```

QY 13 ABAKINMLRITGRPRDGYHLQTVFQMLDLCMTFFHPVDGRVTLNPNISGVE-BODD 71
DB 23 ABAKINMLRITGRPRDGYHLQTVFQMLDLCMTFFHPVDGRVTLNPNISGVE-BODD 82
QY 72 LTVRAANILKSHTGCVRCVCIIDIEKQLPMGGAGGGSSDAATTLVNLRLMGLSKREL 131
DB 72 LTVRAANILKSHTGCVRCVCIIDIEKQLPMGGAGGGSSDAATTLVNLRLMGLSKREL 142
QY 83 LAYRAADIMERRNINRGVTTISIDKQIPVAGLAGSADAAATRGRLFLGQSIDAL 142
DB 132 MDLGRLAGADVPVPGCSAMGEGVEDLOAI-TLPGQVYIIKPDCHVNTGEIFSAENL 190
QY 143 AALGIQIGTDIPPCITNQTAVCTGREGQVTEFLRPPSAWVLLKPNIGISSPVFKALDL 202
DB 191 TRNSAV-----VTMSDF--LAGDNRNDCSEVVCCKLYRPYKCAIDALLCY-ABARL-7 238
QY 203 TBEHIVNEKCKQALENDYHLNLSNRLLEVSNMHHDIKKIKNTMLQCGADGALMS 262
QY 239 GTGACVF 245
DB 263 GSGPTVY 269

```

RESULT 10

AAE31685
 ID AAE31685 standard; Protein; 311 AA.

AC AAE31685;

DT 07-MAR-2003 (first entry)

DE Rhodococcus erythropolis ygbP (ispB) protein.

KM Isoprenoid biosynthesis; enzyme; ygbP; ispB protein.

OS Rhodococcus erythropolis.

PN W0200286094-A2.

PD 31-OCT-2002.

PF 24-APR-2002; 2002MO-US15033.

PR 24-APR-2001; 2001US-285910P.

PA (DUPO) DU PONT DE NEMOURS & CO B I.

PI Biamucci MG, Brzostowicz PC, Cheng Q, Kostichka KM, Nagarajan V;

PI Rouviere PE, Tao L, Thomas SM;

DR MPI; 2003-067638/06.

DR N-PSDB; AAD48796.

XX New nucleic acid molecule from Rhodococcus erythropolis AN12 strain,
 PT useful for obtaining nucleic acid encoding isoprenoid compound
 PT biosynthetic enzyme, and for producing gene products having enhanced or
 PT altered activity -
 PS Claim 4, Page 72-73; 85pp; English.

XX The present invention relates to novel *Rhodococcus erythropolis* AN12
 CC strain isoprenoid biosynthetic enzymes and polynucleotides encoding
 CC such proteins. Sequences of the invention are useful for regulating
 CC isoprenoid biosynthesis in an organism. They are useful for producing
 CC isoprenoid compounds and gene products having enhanced or altered
 CC activity. The present sequence is *Rhodococcus erythropolis* y59p
 CC (isPB) protein.

SQ Sequence 311 AA;

Query Match 18.3%; Score 278; DB 24; Length 311;
 Best Local Similarity 44.1%; Pred. No. 2, 8e-21;

Matches 67; Conservative 22; Mismatches 57; Indels 6; Gaps 3;

QY 13 ABAKIMLMRLTGRREDGYHLQTVFQMLDLCWLTFFHPDDGRVTLR---NPISGVP-E 68
 DB 13 AASKVNLHLAVGDLRDDGYHLLTVGQALSLADTVANAD--TLTVRYIGDDAAAVPTD 70

QY 69 QDDLTFRANLTKSHRGVCVRCIDIERKLPWGGGLGGSSDAATTLVYANRLMGLGSK 128
 DB 71 RNLVWRAEMLAEEGVAPNVEIIEKGI PVAGGAGGSDAAMALVALNSLMGLDPSR 130

QY 129 RELMDLGLRLGADVPVFFVCGSAGGVSSED 160
 DB 131 PDLDAFRAALGSDVPSLHGCTALGTGRGEQ 162

RESULT 11

AA68287
 ID AAB68287 standard; Protein; 401 AA.

AC AAB68287;

DT 09-JUL-2001 (first entry)

XX Amino acid sequence of an isopentenyl monophosphatase kinase.

XX Isopentenyl monophosphatase kinase; IPK; isoprenoid biosynthesis; pigment;
 KM vitamin; essential oil; pathogen resistance; antibiotic; herbicide;
 KM antimutational.

XX Lycopersicon esculentum.

PN M0200132907-A1.

PD 10-MAY-2001.

PF 02-NOV-2000; 2000WO-US30289.

PR 04-NOV-1999; 99US-0434774.

PA (UNIW) UNIV WASHINGTON STATE RES FOUND.

PI Croteau RB, Lange BM;

DR WPI: 2001-308747/32.

DR N-PSDB; AAF85082.

XX New nucleic acid encoding isopentenyl monophosphatase kinase, useful e.g.
 PT for making transgenic plants with increased synthesis of isoprenoids,
 PT e.g. essential oils

PS Example 8; Page 52-53; 62pp; English.

XX The present sequence represents an isopentenyl monophosphatase kinase (IPK)
 CC polypeptide. The enzyme is an isoprenoid biosynthesis stimulator. The
 CC IPK polynucleotide is useful recombinant production of IPK, as a source
 CC of probes, primers and antisense sequences and for increasing/reducing
 CC expression levels of IPK in cells, particularly of essential oil
 CC plants, so as to increase flow through the isoprenoid biosynthesis
 CC pathway, resulting in increased production of e.g. pigments, vitamins
 CC and essential oils, also to increase resistance to pests and pathogens

CC and to improve plant strength. Mutant forms of the IPK polynucleotide
 CC can be used to express forms of IPK that are resistant to IPK-targeted
 CC herbicides, and recombinant IPK can be used to screen for antibiotics,
 CC herbicides and antimutational agents directed against IPK.

SQ Sequence 401 AA;

Query Match 18.3%; Score 277; DB 22; Length 401;
 Best Local Similarity 34.0%; Pred. No. 5, 1e-21;

Matches 71; Conservative 43; Mismatches 89; Indels 6; Gaps 4;

QY 13 ABAKIMLMRLTGRPDGYHLQTVFQMLDLCWLTFFHPDDGRVTLRNPISGVP-EQD 70
 DB 93 SPCKINFLRTSKRDXGDIHDLASLFHVISLGDKIKSLSPSKSKDRISTVAVGPIDER 152

QY 71 DLTFRANLTKSHRGVCVRCIDIERKLPWGGGLGGSSDAATTLVYANRLMGLGSKRE 130
 DB 153 NLTKALNLRKTKGTQNTFYHIDKVPFGAGLGGSSNAATTLMAANQSGCAVATEKE 212

QY 131 LMDLGLRLGADVPVFFVCGSAGGVSSEDLAATLTPQW---PVIKEDCHVNTGEIFSA 187
 DB 213 LQWSEIGSDIPFPFSGHAAYCTGRGEVQDIPSPIDIPVTLIKQOACSTAEVYKR 272

QY 188 ENLFRNSAVVTMSDFLAGDNNDGSEVVC 216
 DB 273 FQDLSSKVPPLS-TLEKISTGIGSDVC 300

RESULT 12

AA68288
 ID AAB68288 standard; Protein; 383 AA.

AC AAB68288;

DT 09-JUL-2001 (first entry)

XX Amino acid sequence of an isopentenyl monophosphatase kinase.

XX Isopentenyl monophosphatase kinase; IPK; isoprenoid biosynthesis; pigment;
 KM vitamin; essential oil; pathogen resistance; antibiotic; herbicide;
 KM antimutational.

XX Arabidopsis thaliana.

PN M0200132907-A1.

PD 10-MAY-2001.

PF 02-NOV-2000; 2000WO-US30289.

PR 04-NOV-1999; 99US-0434774.

PA (UNIW) UNIV WASHINGTON STATE RES FOUND.

PI Croteau RB, Lange BM;

DR WPI: 2001-308747/32.

DR N-PSDB; AAF85083.

XX New nucleic acid encoding isopentenyl monophosphatase kinase, useful e.g.
 PT for making transgenic plants with increased synthesis of isoprenoids,
 PT e.g. essential oils

PS Example 8; Page 56-57; 62pp; English.

XX The present sequence represents an isopentenyl monophosphatase kinase (IPK)
 CC polypeptide. The enzyme is an isoprenoid biosynthesis stimulator. The
 CC IPK polynucleotide is useful recombinant production of IPK, as a source
 CC of probes, primers and antisense sequences and for increasing/reducing
 CC expression levels of IPK in cells, particularly of essential oil
 CC plants, so as to increase flow through the isoprenoid biosynthesis
 CC pathway, resulting in increased production of e.g. pigments, vitamins
 CC and essential oils, also to increase resistance to pests and pathogens

CC and to improve plant strength. Mutant forms of the IPK polynucleotide
CC can be used to express forms of IPK that are resistant to IPK-targeted
CC herbicides, and recombinant IPK can be used to screen for antibiotics,
CC herbicides and antimetabolic agents directed against IPK.

Sequence 383 AA;

Query Match 18.1%; Score 275; DB 22; Length 383;
Best Local Similarity 33.5%; Pred. No. 7,9e-21;
Matches 70; Conservative 42; Mismatches 91; Indels 6; Gaps 4;

QY 13 AAKNLMRLITGRPDGTHLLQTVFQMLDCLMTLTP--PVDDGRVTL-RNPISGVPEQ 70
DB 77 SCKINVFRTITGKRDGPHDLASLFHVISLGDITKPSISPKSDRISTVQGVVDDGR 136
QY 71 DLTVRAANLLKSHTCGVCVCIIDIEKNLPMGGGLGGSSDAATTVLVNRLMGLSKRE 130
DB 137 NLIIRKLNLYRKKTGSNRFVHLDKQVPTGAGLGGSSNATLMAANLNGGLVTENE 196
QY 131 LMDLGLRLGADVVPVFGCSAMGEVSEDLQAITLP--EQWVLIKPDCHVNTGEIFSA 187
DB 197 LQDWSELSISDIPFPFSGAAVCTGREGIVQDLPPFPPLDMVLKPREACSTFAVYKR 256
QY 188 ENLTRSAVVTMSDFLAGDNRDCEVVC 216
DB 257 LRLDQTSNINPLT-LLENVTSKGVSGSTC 284

RESULT 13

AAB68291
ID AAB68291 standard; Protein; 315 AA.

09-JUL-2001 (first entry)

Amino acid sequence of an isopentenyl monophosphatase kinase.

Isopentenyl monophosphatase kinase; IPK; isoprenoid biosynthesis; pigment;
vitamin; essential oil; pathogen resistance; antibiotic; herbicide;
antimicrobial.

Synchocystis sp.

MO200132907-A1.

10-MAY-2001.

02-NOV-2000; 2000MO-US30289.

04-NOV-1999; 99US-0434774.

(UNITV) UNITV WASHINGTON STATE RES FOUND.

Crocteanu RB, Lange EM;

WPI; 2001-308747/32.

N-PSDB; AAF85084.

New nucleic acid encoding isopentenyl monophosphatase kinase, useful e.g.
for making transgenic plants with increased synthesis of isoprenoids,
e.g. essential oils

Example 8; Page 60-61; 62pp; English.

The present sequence represents an isopentenyl monophosphatase kinase (IPK)
polypeptide. The enzyme is an isoprenoid biosynthesis stimulator. The
IPK polynucleotide is useful recombinant production of IPK, as a source
of probes, primers and antisense sequences and for increasing/reducing
expression levels of IPK in cells, particularly of essential oil
plants, so as to increase flow through the isoprenoid biosynthesis
pathway, resulting in increased production of e.g. pigments, vitamins
and essential oils, also to increase resistance to pests and pathogens

CC and to improve plant strength. Mutant forms of the IPK polynucleotide
CC can be used to express forms of IPK that are resistant to IPK-targeted
CC herbicides, and recombinant IPK can be used to screen for antibiotics,
CC herbicides and antimetabolic agents directed against IPK.

Sequence 315 AA;

Query Match 17.8%; Score 270.5; DB 22; Length 315;
Best Local Similarity 30.3%; Pred. No. 1.8e-20;
Matches 91; Conservative 49; Mismatches 109; Indels 51; Gaps 12;

QY 13 AAKNLMRLITGRPDGTHLLQTVFQMLDCLMTLTP--HPVDDGRVTL-RNPISGVPEQ 69
DB 8 AFAKINFLLEIGHRPPGDFHLLVWLOSIALGDKITVRANCTDIRLSCGDSPLAN--DA 65
QY 70 DLTVRAANLLKSHTCGVCVCIIDIEKNLPMGGGLGGSSDAATTVLVNRLMGL 124
DB 66 TNLVRAANLLKSHTCGVCVCIIDIEKNLPMGGGLGGSSDAATTVLVNRLMGL 124
QY 125 GLSKRELMDLGLRLGADVVPVFGCSAMGEVSEDLQAITLP--WVLIK-PDCHVN 180
DB 125 GLTRPELEQLAQLGSDIPFCIGGTAIATGRGILD--PLPDGNCVAVLAKRSIEVS 182
QY 181 T-----GEIFAEINTRSAVVTM-----SDPLAGDNRDCEVVC 216
DB 183 TPMAVQYTRQKFGKXVYNDQSQARRKTHAGPLGIGIQRNPGQIASHIHNDLEKVL 242
QY 217 KLVPRVDAIDALCYA--EARLTGACVPAQPCNREDALEGLKDR-----WL 266
DB 243 PAHQPVAGLRQVLQSAGSLGTPWMSGSPSVYTLTREQAEMAVQVLAIAKEXLNDPVDVWL 302

RESULT 14

AAB70486
ID AAB70486 standard; Protein; 316 AA.

04-MAY-2001 (first entry)

A. thaliana ychb protein without leader sequence.

Isoprenoid biosynthesis; 2C-methyl-D-erythritol-4-phosphate; herbicide;
genetic marker; plant; isoprenoid biosynthesis inhibitor.

Arabidopsis thaliana.

WO200111055-A1.

15-FEB-2001.

03-AUG-2000; 2000MO-EP07548.

04-AUG-1999; 99DE-1036663.

21-SEP-1999; 99DE-1045174.

21-SEP-1999; 99DE-1045175.

11-OCT-1999; 99DE-1048887.

05-NOV-1999; 99DE-1053309.

28-APR-2000; 2000DE-1020996.

(BACH/) BACHER A.

(ZENK/) ZENK M.

Bacher A, Zenk M, Eisenreich W, Fellermeier M, Fischer M, Hecht S,

Herz S, Kis K, Luettgen H, Rohdich F, Sagner S, Schuhr CA;
Kungsintaweekul J;
WPI; 2001-202774/20.
N-PSDB; AAF63675.
Enzymes which operate in the alternative isoprenoid pathway downstream
from 2C-methyl-D-erythritol-4-phosphate, useful for screening a
chemical library for inhibitors of the biosynthesis of isoprenoids -

XX Example 12a; Annex Ea; 194pp; English.

XX The present invention describes functional enzymes (I) which operate in
 CC the alternative isoprenoid pathway downstream from
 CC 2C-methyl-D-erythritol-4-phosphate. (I) can be used as isoprenoid
 CC biosynthesis inhibitor. The enzymes are useful for screening a chemical
 CC library for inhibitors of the biosynthesis of isoprenoids.
 CC 4-diphosphocytidyl-2C-methyl-D-erythritol,
 CC 2C-methyl-D-erythritol-2,4-cyclopyrophosphate or
 CC 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate, or their salts
 CC are useful for screening for inhibitors of the biosynthesis of
 CC isoprenoids. Herbicide resistant enzymes can be used as genetic markers
 CC in any cell that is normally sensitive to the inhibitory effects of
 CC herbicide formation. Methods from the present invention can be used to
 CC produce herbicide resistant enzyme variants which can be incorporated
 CC into plants to confer selective herbicide resistant on the plants. The
 CC methods are also useful for controlling weed by cultivating crops
 CC containing herbicide-resistant genes in the presence of weed-controlling
 CC effective amounts of herbicides. The inhibitors are used for inhibiting
 CC the biosynthesis of isoprenoids in plants, bacteria or protozoa. The
 CC present sequence represents the Arabidopsis thaliana ychb protein without
 CC leader sequence, which is given in an example from the present invention.

XX Sequence 316 AA;

Query Match 17.7%; Score 268; DB 22; Length 316;
 Best Local Similarity 33.0%; Pred. No. 3.5e-20;
 Matches 69; Conservative 42; Mismatches 92; Indels 6; Gaps 4;

QY 13 ABAKNTMLRITGRPPDGYHLLQTVFQMLDCMWTFR-EVDDGRVTEKNPISGVP-EQD 70
 DB 10 SPCKINVFRLITGRKEDGFHDLAFHVISLGTIKESLSPSKSDRLSTWQGVPPVGR 69
 QY 71 DITVRAANILKSHRGVRCVCDIEKNLPMGGGAGGSSDATTLVNRLMGLSKRE 130
 DB 70 NLIIRALNLYRKKTGSNRFPMIHLDKVPTGAGLGGSSNATLMAANLNGGLVTENE 129
 QY 131 LMDLGLRLGADVVPVFGCSAMGEGVSEDLQATLP--EQWFIYIKPDCHVNTGEISFA 187
 DB 130 LQWMSSEIGSDIPFFSHGAAVCTGRGEIVQDLPPFPPLDPMWLIKRBACTALEVKR 189
 QY 188 ENUTRNSAVVTMSDFLAGDNNDCEVVC 216
 DB 190 ERLDQTSNINPLT-LKXVTSNGVSQISIC 217

RESULT 15
 AAB70487
 ID AAB70487 standard; Protein; 383 AA.
 AC AAB70487;
 XX
 XX 04-MAY-2001 (first entry)
 DT Arabidopsis thaliana ychb protein.
 DE Arabidopsis thaliana ychb protein.
 XX
 XX Isoprenoid biosynthesis; 2C-methyl-D-erythritol-4-phosphate; herbicide;
 KM genetic marker; plant; isoprenoid biosynthesis inhibitor.
 XX
 XX Arabidopsis thaliana.
 OS
 XX
 XX W0200111055-A1.
 PD 15-FEB-2001.
 XX
 XX 03-AUG-2000; 2000WO-EP07548.
 PF
 XX
 XX 04-AUG-1999; 99DE-1036663.
 PR 21-SEP-1999; 99DE-1045174.
 PR 21-SEP-1999; 99DE-1045175.
 PR 11-OCT-1999; 99DE-1048887.
 PR 05-NOV-1999; 99DE-1053309.

PR 28-APR-2000; 2000DE-1020996.
 XX
 XX (BACH/) BACHER A.
 PA (ZENK/) ZENK M.
 XX
 XX Bacher A, Zenk M, Bisenreich W, Fellermeier M, Fischer M, Hecht S;
 PI Herz S, Kis K, Luettgen H, Rohdich F, Sagner S, Schuhr CA;
 PI Wungslintwekul J;
 XX
 XX WPI; 2001-202774/20.
 DR N-PSDB; AAF63680.
 XX
 XX Enzymes which operate in the alternative isoprenoid pathway downstream
 PT from 2C-methyl-D-erythritol-4-phosphate, useful for screening a
 PT chemical library for inhibitors of the biosynthesis of isoprenoids -
 XX
 XX Example 12b; Annex Eb; 194pp; English.

XX The present invention describes functional enzymes (I) which operate in
 CC the alternative isoprenoid pathway downstream from
 CC 2C-methyl-D-erythritol-4-phosphate. (I) can be used as isoprenoid
 CC biosynthesis inhibitor. The enzymes are useful for screening a chemical
 CC library for inhibitors of the biosynthesis of isoprenoids.
 CC 4-diphosphocytidyl-2C-methyl-D-erythritol,
 CC 2C-methyl-D-erythritol-2,4-cyclopyrophosphate or
 CC 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate, or their salts
 CC are useful for screening for inhibitors of the biosynthesis of
 CC isoprenoids. Herbicide resistant enzymes can be used as genetic markers
 CC in any cell that is normally sensitive to the inhibitory effects of
 CC herbicide formation. Methods from the present invention can be used to
 CC produce herbicide resistant enzyme variants which can be incorporated
 CC into plants to confer selective herbicide resistant on the plants. The
 CC methods are also useful for controlling weed by cultivating crops
 CC containing herbicide-resistant genes in the presence of weed-controlling
 CC effective amounts of herbicides. The inhibitors are used for inhibiting
 CC the biosynthesis of isoprenoids in plants, bacteria or protozoa. The
 CC present sequence represents the Arabidopsis thaliana ychb protein,
 CC which is given in an example from the present invention.

XX Sequence 383 AA;

Query Match 17.7%; Score 268; DB 22; Length 383;
 Best Local Similarity 33.0%; Pred. No. 4.6e-20;
 Matches 69; Conservative 42; Mismatches 92; Indels 6; Gaps 4;

QY 13 ABAKNTMLRITGRPPDGYHLLQTVFQMLDCMWTFR-EVDDGRVTEKNPISGVP-EQD 70
 DB 77 SPCKINVFRLITGRKEDGFHDLAFHVISLGTIKESLSPSKSDRLSTWQGVPPVGR 136
 QY 71 DITVRAANILKSHRGVRCVCDIEKNLPMGGGAGGSSDATTLVNRLMGLSKRE 130
 DB 137 NLIIRALNLYRKKTGSNRFPMIHLDKVPTGAGLGGSSNATLMAANLNGGLVTENE 196
 QY 131 LMDLGLRLGADVVPVFGCSAMGEGVSEDLQATLP--EQWFIYIKPDCHVNTGEISFA 187
 DB 197 LQWMSSEIGSDIPFFSHGAAVCTGRGEIVQDLPPFPPLDPMWLIKRBACTALEVKR 256
 QY 188 ENUTRNSAVVTMSDFLAGDNNDCEVVC 216
 DB 257 ERLDQTSNINPLT-LKXVTSNGVSQISIC 284

Search completed: January 29, 2004, 15:49:42
 Job time : 28.0514 secs

; Sequence 6, Application US/09434774A

```

: Patent No. 6235514
:
: GENERAL INFORMATION:
:
: APPLICANT: Croteau, Rodney B
:
: APPLICANT: Lange, Bernd M
:
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ISOPRENTENYL
:
: TITLE OF INVENTION: MONOPHOSPHATE KINASE, AND METHODS OF USE
:
: FILE REFERENCE: wsbu14448
:
: CURRENT APPLICATION NUMBER: US/09/434,774A
:
: CURRENT FILING DATE: 1999-11-04
:
: NUMBER OF SEQ. ID NOS.: 16
:
: SOFTWARE: Patentln Ver. 2.0
:
: SEQ ID NO 6
:
: LENGTH: 283
:
: TYPE: PR1
:
: ORGANISM: Escherichia coli
:
: US-09-434-774-6

```

Qy	10	RWAPAKNLMIAITGRPDYHLLQIVOMDLCDMLTFHVVDGGRATNPLISGYEQ	69
Db	4	QWSPAKNLMFLYITQGRADYHLLQIVOMDLCDMLTFHVVDGGRATNPLISGYEQ	63
Qy	70	DDLTPRANLL-----KSHTKCVRGVCIIDENKLMGGGLGGSSDAATTVLNRLL	122
Db	64	DNIVAPARLLKMTAADSGRLPTG--SGSNISIDRLPMGGGLGGSSDAATTVLNRLL	122
Qy	122	WGLGSKREMDGLRLGADVPLVFGVGGAMGEGVEDLOAITLPEQOFVIRKPDCHYNT	181
Db	122	WGGGLSMDSLAEMGLLKGADVPLVFGVGGAAFAFGEGEILITPDPPEKRYLYAHGVGSILPT	181
Qy	182	GEIFSAENLTRNSAVVTMSDPLAGDNRNDCSEVCKLYRPVKDAIDALLCYAABRLTGTG	241
Db	182	PVTFKQPELPRNTPKRSIRITLLKCEFSNDCVEIARKEFEADVAVISWLLLEAPRLTGTG	241
Qy	242	ACVFAPOCKNEDEASLLEGLKDRML-VFLAKSI:NGSALYKYL	282
Db	242	ACVFAEEDTSEAKQVLEQAPE-VLNGVFAKAGNLSFLHRAH	282

	Query Match	38.9%	Score 589.5;	DB 4:	Length 283;
	Best Local Similarity	43.9%;	Pred. No.	8.2e-57;	
	Matches 119;	Conservative 58;	Mismatches 89;	Indels 5;	Gaps 4
QY	10 RWEPAPAKNTMLNITGRRRPDGYHLLQTVAOMDLCPMFLFHFVDGKRTLTNNPISGVBEQ	69			
Dd	9 RVSPSPAKNLFLHITRRENGTHDELTIQLLDIDVMHTFISSDELDIQSG-LGEVGLTE	67			
QY	70 DDITVRANILKSHTCGVGCVIDIEKNLPMGGGIAGSSSDAATTVIVIANRLMGJLSKR	122			
Dd	68 QNLIYRAAQTLRPQAQNPGGLHIKIENKLPMSGAGIAGSSSNAAATTVIVANQLMQCGLTEE	122			
QY	130 ELMDLGIRLCADVDVPVFQGCSANGGVSIEDLAITLPEDQMFIITIPDCHWTGTGELFSKEN	183			

Db 128 QLAPEVTLGADVPIFFITGLNANNEGISEHHSFLDIDQKQIVALKPDQPSITVOLLFSQKT 187

Qy 190 LTRNSAVVTMSPF-LAGDN-RKDCEVYCKIYRPYKDAIDMLTCYAAKRLTGACVFAQ 247

Db 188 LTRDSPTTFCAYOLEPSPNGNPEPLARELYPEVBEAMQYLDQFGHAKLTGACVFAE 247

Qy 248 PCNKEDAESALGLKDRMLVFLAKINOSAL 278

Db 248 VTDEMAYVDILGHAPCK--AYIYHSLKESPL 276

Qy	1	A P K N I M L R T G R R P D G V T H L Q T F Q C M L D C M U T F P W D G V T L E N P I S G V P - S Q D D	71
		: : : : :	
Db	23	A P K N I F T L D T L F K D D G T H E I E M M T V Y D L N D L S F E K R I D N K I V D I E H N Y P N D N K X	82
		: : : : : :	
Qy	72	L T V P A A N L I K S H T G C V R G V C I D I E K N L P M G G L G G G S D A A T T L V L N L R K W G L S R E L	131
		: : : : : :	
Db	83	L A Y K A D L M F E R F N I N E G V T I S I D D I P V S A G L A G G S A D A A T Y R G L N R L F G L Q S I D A L	142
		: : : : : :	
Qy	132	M D I G L T G L A D V P V F P G C S A M G E V G S E D L Q A I T L - P E Q W F I I K P D C H V N T G E I F S A E N L	190
		: : : : : :	
Db	143	A A L I G I D I G D I F P C I Y N O T A V C T G E G E O V T T L K R P P S A M V Y L A P N I G I S S P D V F K A D L	202
		: : : : : :	
Qy	191	T R N S A V - - - - - V T M S D F - L A G D N N D C S E V V C K L Y R P V K A I D A L L - C V A E - A E L T	238
		: : : : : :	
Db	203	T E H H I Y H N E K C K Q A L E N N D Y H L L C N S L N B R L E P V S M A M G H P I D I K K I N M L Q C G A D G A L M S	262
		: : : : : :	
Qy	239	G T G A C V F	245
		:	
Db	263	G S G P T V Y	269
		:	

CURRENT FILING DATE: 1999-11-04
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 16
LENGTH: 315
TYPE: PRT
ORGANISM: *Synechocystis* sp.
US-09-434-774-16

Query Match 17.8%; Score 270.5; DB 3; Length 315;
Best Local Similarity 30.3%; Pred. No. 1.7e-21;
Matches 91; Conservative 49; Mismatches 109; Indels 51; Gaps 12;

13 A P A K N L M L R I T G R R D G Y H L L Q T V F O M L D C M L T F - H P V D D G V T L R N P I S G V P E Q 69
8 A P K I N L F E I L A D R P D G S H E L V M L Q S I A L G K I T V R A N G T D I R L S G D S P L A N - D A 65
70 D D I T V R A N I L - - - - - K S H T G C V R G C I D I E K N L P M G G L A G G S S D A T T L V L N R L M G L 124
66 T N L A Y R A Q L M N N F P Q A I D N - Y G G V D I T L T K H I P M A G L A G S S A D A A V L G D L L M N L 124
125 G L S K R I M L G L R L G A D V P V P F G C S A M G S G S E D L Q A I T L P E Q - - W F V I T K - P D C H V N 180
125 G L T R P E L Q L A A Q L G S D I P F C I G G G T A I A T G R G E I L D - P L P D G M C F W V V L A K R S I E V S 182
181 T - - - - - G E I S A E N L T R N S A V T M - - - - - S P L A G D N R N D C S E V V C 216
183 T P M A Y O T R O K R K N V I N D Q S R A R K T I H A G P L L O G I O H N P Q I A S H I N D L K R V L 242
217 K L Y R P V K A I D A L L C Y A - E A R I T G T A C Y P A O F C N K E D A E S A L E G L K D R - - - - - W L 266
243 P A H Q V A O L R O V L O S A G L G T W M S G S G P S V F T L C R E Q A E A V O L A I A K E K L A N D P V D F A L 302

RESULT 9

US-09-434-774-2
Sequence 2, Application US/09434774A
Patent No. 6235514
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Lange, Bernd M
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ISOPRENTAL
FILE REFERENCE: wslut14448
CURRENT APPLICATION NUMBER: US/09/434,774A
CURRENT FILING DATE: 1999-11-04
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 2
LENGTH: 405
TYPE: PRT
ORGANISM: *Mentha piperita*
US-09-434-774-2

Query Match 17.3%; Score 263; DB 3; Length 405;
Best Local Similarity 30.2%; Pred. No. 1.6e-20;
Matches 76; Conservative 50; Mismatches 104; Indels 22; Gaps 7;

13 A P A K N L M L R I T G R R D G Y H L L Q T V F O M L D C M L T F - H P V D D G V T L R N P I S G V P E Q 69
102 S P C K I N V F R I T G K R E D G F H D L A S L F H V I S L D K I K F S L S P K F N S F V T N - V P G V P L D S 160
70 D D I T V R A N I L K S H T G C V R G C I D I E K N L P M G G L A G G S S D A T T L V L N R L M G L S K R 129
161 K O E I I T A L N L F R K K T D T H F M I H L D K V P T A G L A G G S S N A T A L M A A N O F S G C I A T E K 220
130 E I M D L G L R L G A D V P V P F G C S A M G S G S E D L Q A I T L P - - E O M F Y I I K P D C H V N T G E I F S 186
221 D L O E M S G E I G S D I P F F F S G A A Y C T G R G E V E D I P P V P R O L S V T L M K P Q E A C P F G E Y K 280
187 A E N I T R N S A V T M - - - - - S P L A G D N R N D C S E V V C K L Y R P V K A I D A L - - L C Y 232
281 R I L L O T S I D I P V I L L E K I S K G I S O D V C N D I B P A P A E V V S L K - L K O R I A A G R S O Y 339

QY 233 A E A R I T G T A C V 244
DB 340 D A V F M S G S G S T I 351

RESULT 10

US-09-198-452A-1236
Sequence 1236, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffla, R.
TITLE OF INVENTION: *Chlamydia pneumoniae* genomic sequence and polypeptides, fragments
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 1236
LENGTH: 141
TYPE: PRT
ORGANISM: *Chlamydia pneumoniae*
US-09-198-452A-1236

Query Match 11.7%; Score 178; DB 4; Length 141;
Best Local Similarity 32.3%; Pred. No. 8.4e-12;
Matches 43; Conservative 30; Mismatches 58; Indels 2; Gaps 1;

QY 13 A P A K N L M L R I T G R R D G Y H L L Q T V F O M L D C M L T F H P V D D G V T L R N P I S G V P E Q D D L 72
DB 5 S P A K I N L F E I L A D R P D G S H E L V M L Q S I A L G K I T V R A N G T D I R L S G D S P L A N - D A 65
73 T V R A N I L K S H T G C V R G C I D I E K N L P M G G L A G G S S D A T T L V L N R L M G L S K R E I M 132
63 I W K S I B I R R E T O H O V S W H N K S I F L O S G L G G S S N A T A L Y A L N E H P O T H I P I T L Q 122
QY 133 D L G L R L G A D V P V F 145
DB 123 L M A R B I S D V P P F 135

RESULT 11

US-09-634-238-305
Sequence 305, Application US/09634238
Patent No. 6544772
GENERAL INFORMATION:
APPLICANT: Glenn, Matthew
APPLICANT: Hayukala, Ilka J.
APPLICANT: Bloksberg, Leonard, N.
APPLICANT: Lubbers, Mark W.
APPLICANT: Dekker, James
APPLICANT: Christensen, Anna C.
APPLICANT: Holland, Ross
APPLICANT: O'Toole, Paul W.
APPLICANT: Reid, Julian R.
APPLICANT: Coolbear, Timothy
TITLE OF INVENTION: Polynucleotides, materials incorporating
TITLE OF INVENTION: them and methods for using them.
FILE REFERENCE: 11000.104301
CURRENT APPLICATION NUMBER: US/09/634,238
CURRENT FILING DATE: 2000-08-08
NUMBER OF SEQ ID NOS: 422
SOFTWARE: PatsSeq for Windows Version 4.0
SEQ ID NO 305
LENGTH: 99
TYPE: PRT
ORGANISM: *Lactobacillus rhamnosus*
US-09-634-238-305

Query Match 10.6%; Score 160; DB 4; Length 99;
Best Local Similarity 44.4%; Pred. No. 4.9e-10;
Matches 36; Conservative 15; Mismatches 30; Indels 0; Gaps 0;

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;; FILE REFERENCE: GTC-007
;; CURRENT APPLICATION NUMBER: US/09/134,001C
;; CURRENT FILING DATE: 1998-08-13
;; PRIOR APPLICATION NUMBER: US 60/064,964
;; PRIOR FILING DATE: 1997-11-08
;; PRIOR APPLICATION NUMBER: US 60/055,779
;; PRIOR FILING DATE: 1997-08-14
;; NUMBER OF SEQ ID NOS: 5674
;; SEQ ID NO 4141
;; LENGTH: 315
;; TYPE: PRT
;; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4141

Job time : 9.94468 secs

Query Match 5.9%; Score 89; DB 4; Length 315;
Best Local Similarity 20.0%; Pred. No. 0.19;
Matches 30; Conservative 38; Mismatches 66; Indels 16; Gaps 5;

QY 63 ISGVPEQDD-LTVRANILKSHTCVRCVACIDIEKNLPMGGIGGSSDAATLIVNLR 120
DB 61 LELPFDENNYIYQRLNARKYNTLPSQLIMESDIPLANGSSASALVGLPIANY 120
QY 121 LMGGLSKRELMDGLRLGA--DPEVFYFG--CSAMGEVSE--DLQAITLPEQWFI 173
DB 121 FGNIGSKRELQGLATEIGHNPVAPITYGGLIAGFYNPITKITVARIIEVPHVDIILE 180
QY 174 KPDCHVNT-----GEIISAENLTNSAV 196
DB 181 IPEYELRTEDSRRLVLPDTFSHKGVQNSAI 210

RESULT 15

US-09-252-991A-24594
; Sequence 24594; Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24594
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24594

Query Match 5.7%; Score 86; DB 4; Length 312;
Best Local Similarity 30.8%; Pred. No. 0.39;
Matches 44; Conservative 15; Mismatches 58; Indels 26; Gaps 7;

QY 50 HPYDGRVTLRNPIS-GVPEQDDLTVRANILKSHTCVRCVACIDIEKNLPMGGIGGGS 108
DB 155 HPQCBGA-----PLSGIIPRILFVTLVNIHPEGMAGVGSAGLDEANGLAGIAL 209
QY 109 SDAITLVNLRMLMGGLSKRELMDGLGLGADVPEVFYFGCSAMGEVSEDLQAITLPEQ 168
DB 210 QVPEGLVALVLAGVSKRFKAMLVGAASGLVEPLFAVLIC-AMLVGLS---ALLP-- 262
QY 169 W-----FVI---IKPDCH 178
DB 263 WGLAAAGAMLFVVTETIILPESH 285

Search completed: January 29, 2004, 15:57:05

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 : Search time 19.4547 Seconds
(without alignments)
3044.503 Million cell updates/sec

Title: US-09-941-947a-12
Perfect score: 1516
Sequence: 1 MDYAAGCGERPAPAKLMLK.....LVFLAKGINSALYKLEEG 285

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*

12: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*

13: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*

14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*

15: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep:*

16: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

17: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1516	100.0	285	US-09-934-903-8	Sequence 8, Appli
2	1516	100.0	285	US-09-934-868-74	Sequence 74, Appli
3	1516	100.0	285	US-09-941-947a-12	Sequence 12, Appli
4	638	42.1	279	US-10-369-493-13766	Sequence 13766, A
5	620	40.9	279	US-10-369-493-301	Sequence 201, App
6	616	40.6	283	US-10-369-493-802	Sequence 802, App
7	584.5	38.6	280	US-10-369-493-21079	Sequence 21079, A
8	584	38.5	294	US-10-369-493-15324	Sequence 15324, A
9	582	38.4	276	US-10-369-493-15690	Sequence 15690, A
10	582	38.4	276	US-10-369-493-16082	Sequence 16082, A
11	580.5	38.3	298	US-10-369-493-17751	Sequence 17751, A
12	572	37.7	270	US-10-369-493-9184	Sequence 9184, Ap
13	572	37.7	270	US-10-369-493-9359	Sequence 9359, Ap
14	556	36.7	213	US-10-369-493-19675	Sequence 19675, A
15	533	35.2	233	US-10-369-493-8419	Sequence 8419, Ap

16	407	26.8	284	US-10-369-493-9956	Sequence 9956, Ap
17	341	22.5	288	US-10-369-493-16541	Sequence 16541, A
18	327	21.6	269	US-10-369-493-7851	Sequence 7851, Ap
19	324	21.4	289	US-10-369-493-23046	Sequence 23046, A
20	320	21.1	287	US-10-369-493-9660	Sequence 9660, Ap
21	317	20.9	287	US-10-369-493-17097	Sequence 17097, A
22	292	19.3	316	US-10-369-493-20265	Sequence 20265, A
23	286.5	18.9	295	US-10-369-493-12283	Sequence 12283, A
24	281	18.5	297	US-10-369-493-11123	Sequence 11123, A
25	278	18.3	311	US-10-369-493-11241	Sequence 11241, A
26	274	18.1	292	US-10-369-493-11764	Sequence 11764, A
27	273.5	18.0	261	US-10-369-493-14419	Sequence 14419, A
28	273.5	18.0	261	US-10-369-493-14670	Sequence 14670, A
29	273.5	18.0	261	US-10-369-493-15150	Sequence 15150, A
30	273.5	18.0	261	US-10-369-493-15150	Sequence 15150, A
31	270.5	17.8	315	US-10-369-493-22607	Sequence 22607, Ap
32	267	17.6	306	US-09-112-363-186	Sequence 186, App
33	254	16.8	240	US-10-369-493-17829	Sequence 17829, A
34	250.5	16.5	268	US-10-369-493-65	Sequence 65, Appli
35	239	15.8	295	US-10-369-493-16874	Sequence 16874, A
36	233.5	15.4	271	US-10-369-493-3024	Sequence 3024, Ap
37	216	14.2	311	US-09-738-626-4515	Sequence 4515, Ap
38	203	13.4	245	US-10-369-493-10645	Sequence 10645, A
39	178	11.7	141	US-10-264-227-358	Sequence 1236, Ap
40	160	10.6	99	US-10-264-213-208	Sequence 128, App
41	127	8.4	238	US-09-882-227-358	Sequence 358, App
42	125	8.2	314	US-09-815-242-10963	Sequence 10963, A
43	116	7.7	575	US-10-104-047-2815	Sequence 2815, Ap
44	116	7.7	813	US-10-108-2604-3901	Sequence 3901, Ap
45	116	7.7	990	US-10-094-749-2568	Sequence 2568, Ap

ALIGNMENTS

RESULT 1

US-09-934-903-8

Sequence 8, Application US/09934903

Patent No. US20020102690A1

GENERAL INFORMATION:

APPLICANT: Koffas, Mattheos

APPLICANT: Odono, J. Martin

APPLICANT: Schenle, Andreas J.

APPLICANT: Tomb, Jean-Francois

APPLICANT: Rouviere, Pierre

APPLICANT: Picataggio, Stephen

APPLICANT: Cheng, Qiong

TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production

FILE REFERENCE: C1646 US NA

CURRENT APPLICATION NUMBER: US/09/934, 903

CURRENT FILING DATE: 2001-08-22

PRIOR APPLICATION NUMBER: 60/229, 907

PRIOR FILING DATE: September 1, 2001

NUMBER OF SEQ ID NOS: 24

SOFTWARE: Microsoft Office 97

SEQ ID NO 8

LENGTH: 285

TYPE: PRT

ORGANISM: Methylobionas 16a

FEATURE:

OTHER INFORMATION: Amino acid sequences encoded by ORF4

US-09-934-903-8

Query Match 100.0%; Score 1516; DB 10; Length 285;

Best Local Similarity 100.0%; Pred. No. 5.7e-156; Indels 0; Gaps 0;

Matches 285; Conservative 0; Mismatches 0;

Cy 1 MDYAAGCGERPAPAKLMLKRTGRPDYHLLQTVFQMLDCLDWLTFHVDGRTLR 60

Db 1 MDYAAGCGERPAPAKLMLKRTGRPDYHLLQTVFQMLDCLDWLTFHVDGRTLR 60

Cy 61 NPISGVEQDDLTFRANLTKSHTGCVRGVCIDIEKMLPGGGIGGSSDAATLVVILNR 120

Db	61	NPISGVPEGGDILTVRAANLTKSHTECVRGVCIDIERKLPMGGGLGGSSDAATTLTVLNR	120
Qy	121	LMWIGSKREIMDLGRLGADVPVPLFGSSANGESVSDELQATLPEPMFVITIPDCHN	180
Db	121	LMWIGSKREIMDLGRLGADVPVPLFGSSANGESVSDELQATLPEPMFVITIPDCHN	180
Qy	181	TGSIIPSAENLTENSAVVTMSDFLAGDNRDCEVVCKLYRPVKAIDALLCYAEARLTGT	240
Db	181	TGSIIPSAENLTENSAVVTMSDFLAGDNRDCEVVCKLYRPVKAIDALLCYAEARLTGT	240
Qy	241	GACVFAQFCMKEDAESALGELKDRMTVFLAKLINSALYKYLEGQ	285
Db	241	GACVFAQFCMKEDAESALGELKDRMTVFLAKLINSALYKYLEGQ	285

RESULT2
US-09-934-868-74
; Sequence 74, Application US/09934868
; Patent No. US20020137190A1
; GENERAL INFORMATION:

```

? APPLICANT: Kottas, Mattheos
? APPLICANT: Odom, James M
? APPLICANT: Schenzle, Andreas J
? TITLE OF INVENTION: IDENTIFYING METHANOTROPHIC BACTERIAL STRAIN
? FILE REFERENCE: C14596 US NA
? CURRENT APPLICATION NUMBER: US/09/934, 868
? CURRENT FILING DATE: 2001-08-22
? PRIOR APPLICATION NUMBER: 60/229, 858
? PRIOR FILING DATE: 2000-09-01
? NUMBER OF SEQ ID NOS: 81
? SOFTWARE: Microsoft Office 97
? SEQ ID NO 74
? LENGTH: 285
? TYPE: PRT
? ORGANISM: Methylomonas 16a
? FEATURE:
? OTHER INFORMATION: Amino acid sequences encoded by ISPE
? US-09-934-868-74

```

Query Match	100.0%	Score 1516;	DB 10;	Length 285;
Best Local Similarity	100.0%	Pred. No. 5.7e-156;		
Matches 285;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	MDYAAAGGSEWMPAPALINMLRITGSRPGVGLLTQVQMDLDCDMLFPHVDGGRVTLR	60
Db	1	MDYAAAGGSEWMPAPALINMLRITGSRPGVGLLTQVQMDLDCDMLFPHVDGGRVTLR	60
QY	61	NPIISGVPEDDDLVRRAANLKSHTCGVGVGVCIDIKKLPMSGGIAGGGSDDAATLVVLNR	120
Db	61	NPIISGVPEDDDLVRRAANLKSHTCGVGVGVCIDIKKLPMSGGIAGGGSDDAATLVVLNR	120
QY	121	LMGLGSKSEKIMDLGRLGADVPVYFVFGSSAMGSEVSESDAATLPPEQWPIYIRPDCHN	180
Db	121	LMGLGSKSEKIMDLGRLGADVPVYFVFGSSAMGSEVSESDAATLPPEQWPIYIRPDCHN	180
QY	181	TGEISFAEMLTRNSAAVYMSDFLAGDNRRDCEVVCCKYRPVVKALIDALLCYAARLTGT	240
Db	181	TGEISFAEMLTRNSAAVYMSDFLAGDNRRDCEVVCCKYRPVVKALIDALLCYAARLTGT	240
QY	241	GACVFAQPCNKEDAESALEGLKDRMLVPLAKINLSALYKKEEQ	285
Db	241	GACVFAQPCNKEDAESALEGLKDRMLVPLAKINLSALYKKEEQ	285

RESULT 3
 US-09-941-947A-12
 ; Sequence 12, Application US/0991947A
 ; Publication No. US20030003528A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Broszowiec, Patricia C.
 ; APPLICANT: Cheng, Qiong
 ; APPLICANT: DiCosimo, Deana J.

```

1  APPLICANT: Koffas, Mattheos
2  APPLICANT: Miller, Edward S. Jr.
3  APPLICANT: Odum, J. Martin
4  APPLICANT: Picataggio, Steve
5  APPLICANT: Rouviere, Pierre E.
6  TITLE OF INVENTION: CHAOTICNOID PRODUCTION FROM A SINGLE CARBON SOURCE
7  FILE REFERENCE: C11903 US NA
8  CURRENT APPLICATION NUMBER: US/09/941,947A
9  CURRENT FILING DATE: 2001-09-01
10 PRIOR APPLICATION NUMBER: 60/229,907
11 PRIOR FILING DATE: 2000-09-01
12 PRIOR APPLICATION NUMBER: 60/229,858
13 PRIOR FILING DATE: 2000-09-01
14 NUMBER OF SEQ ID NOS: 60
15 SOFTWARE: Microsoft Office 97
16 SEQ ID NO 12
17 LENGTH: 285
18 TYPE: PRT
19 ORGANISM: Methylobionas 16a
20 US-05-941-947A-12

```

Query Match	100.0%	Score 1516	DB 11	Length 285
Best Local Similarity	100.0%	Pred. No. 57e-156		
Matches 285	0	Mismatches 0	Indels 0	Gaps 0

QY	1	MDVAAGGSEWMPAPAKNTLMRTTGRSPDGHLLQTFQMLDLCMTLTFRFVDDGRVTLR	60
QY	1	MDVAAGGSEWMPAPAKNTLMRTTGRSPDGHLLQTFQMLDLCMTLTFRFVDDGRVTLR	60
Db	1	MDVAAGGSEWMPAPAKNTLMRTTGRSPDGHLLQTFQMLDLCMTLTFRFVDDGRVTLR	60
QY	61	NPISGVEODDLTVRAANTLLKSHTCVARGVCIDIEKQLPMGGSGICGSSSDAATLLTVLNR	120
Db	61	NPISGVEODDLTVRAANTLLKSHTCVARGVCIDIEKQLPMGGSGICGSSSDAATLLTVLNR	120
QY	121	LMGIGSKRELMQGLRLGADVLPFVFGCSAMGSGVEDLQATILPEQWFIILRPDCGVN	180
Db	121	LMGIGSKRELMQGLRLGADVLPFVFGCSAMGSGVEDLQATILPEQWFIILRPDCGVN	180
QY	181	TGBIFSAENLTRNSAAVVTMSDPLAGDNRNDCSEVVCKLVRPVKCAIDALLCYAARLTGT	240
Db	181	TGBIFSAENLTRNSAAVVTMSDPLAGDNRNDCSEVVCKLVRPVKCAIDALLCYAARLTGT	240
QY	241	GACVPAFOCNKEDAESALBGLKQRMVFLFLAKGNQSLATYKRLBEG	285
Db	241	GACVPAFOCNKEDAESALBGLKQRMVFLFLAKGNQSLATYKRLBEG	285

RESULT 4
US-10-369-493-13766
; Sequence 13766, Application US/10369493
; Publication No. US20030233675A1

GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ. ID NOS: 47374
 SEQ. ID NO 15766
 LENGTH: 279
 TYPE: PRT
 ORGANISM: *Pseudomonas fluorescens*
 US-10-369-493-13766

Query March	42.1%	Score 638;	DB 12;	Length 279;
Best Local Similarity	49.6%	Pred. No. 1	2e-60;	
Matches 135;	Conservative	41;	Mismatches	96;
			Indels	0;
			Gaps	0;

QY 12 PAKINIMLRITGRRPDGYHLLQTVFQMLDLCMLTFHPVDDGAVTLNPISGVPEODD 71
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 201
 ; LENGTH: 279
 ; TYPE: PRT
 ; ORGANISM: Xenorhabdus nematophilus
 ; US-10-369-493-201

Db 5 PEPKINIMLRITGRRPDGYHLLQTVFQMLDLCMLTFHPVDDGAVTLNPISGVPEODD 64
 72 LTVRAANLKSHGTCVRGVCIDIEKNLPMGGGLGGSSNAATTLVLAIRLWELGSKDEL 131
 65 LIVRAAKLQESASCSLGIDIWIDKLPFGGGLGGSSNAATTLGLNHLRLGMBEDRL 124
 132 MDGLRLGADVVPVFGCSAMGEGVSEDLQAITLPEOWFVILKPDCHVNTGEIFSAEMLT 191
 125 AALGTLIGADVVPVFGHAAPAEVGSGKLTPEAPAPVTVVLPQVSVTAEIFSDPLLT 184
 129 RNSAVVMSDFLAGNNDCEVVCCKLTPPVDAIDALLCYAARLTGACVAFQFCNK 251
 185 RNPPLKAPVPEGNSRNDCLPVSSRIPVVRNALNLGNTFAKLTGTGSCVFGGFPSSK 244
 252 EDASALBGLKDRMLVFLAKGNOSALYKLE 283
 245 AEDKVSALTETLTGPFVAKGSNVSILHRKIQ 276

RESULT 5
 US-10-369-493-201
 ; Sequence 201, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 201
 ; LENGTH: 279
 ; TYPE: PRT
 ; ORGANISM: Xenorhabdus nematophilus
 ; US-10-369-493-201

Query Match 40.9%; Score 620; DB 12; Length 279;
 Best Local Similarity 48.0%; Pred. No. 11e-58;
 Matches 132; Conservative 36; Mismatches 101; Indels 6; Gaps 2;

QY 11 WPAPAKINIMLRITGRRPDGYHLLQTVFQMLDLCMLTFHPVDDGAVTLNPISGVPEOD 70
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 201
 ; LENGTH: 279
 ; TYPE: PRT
 ; ORGANISM: Xenorhabdus nematophilus
 ; US-10-369-493-201

Db 5 WSPAPAKINIMLRITGRRPDGYHLLQTVFQMLDLCMLTFHPVDDGAVTLNPISGVPEOD 64
 71 DITVRAANLKSH-----TGC-VRGVCIDIEKNLPMGGGLGGSSDAATTLVLAIRLWEL 124
 65 NITVRAARLLQDHAKRGSTGYRADLHHRRLPMGGGLGGSSNAATTLVLAIRLWEL 124
 125 GLSKRELMDLGLRLGADVVPVFGCSAMGEGVSEDLQAITLPEOWFVILKPDCHVNTGEI 184
 125 NLSDELQGLGTLGADVVPVFGHAAPAEVGSGKLTPEAPAPVTVVLPQVSVTAEIFSDPL 184
 165 FSAENVLTNSAVVMSDFLAGNNDCEVVCCKLTPPVDAIDALLCYAARLTGACVAFQFCNK 244
 185 FTDPLKAPVPEGNSRNDCLPVSSRIPVVRNALNLGNTFAKLTGTGSCVFGGFPSSK 244
 245 PAQFCNKEDASALBGLKDRMLVFLAKGNOSALY 279
 245 PGEFESPPASARKVLAQAPERMKGFTVARGVNTISPLH 279

RESULT 6
 US-10-369-493-802

Sequence 802, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 802
 ; LENGTH: 283
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 ; US-10-369-493-802

Query Match 40.6%; Score 616; DB 12; Length 283;
 Best Local Similarity 47.9%; Pred. No. 3e-58;
 Matches 135; Conservative 39; Mismatches 96; Indels 12; Gaps 4;

QY 10 WPAPAKINIMLRITGRRPDGYHLLQTVFQMLDLCMLTFHPVDDGAVTLNPISGVPEQ 69
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 201
 ; LENGTH: 279
 ; TYPE: PRT
 ; ORGANISM: Xenorhabdus nematophilus
 ; US-10-369-493-201

Db 4 QWSPAPAKINIMLRITGRRPDGYHLLQTVFQMLDLCMLTFHPVDDGAVTLNPISGVPEQ 63
 70 DITVRAANLKSH-----KSHGTCVRGVCIDIEKNLPMGGGLGGSSDAATTLVLAIRL 121
 64 DNLVRAARLLKMAADSGRLFG--SGANISIDKRLPMGGGLGGSSNAATTLVLAIRL 121
 122 WGLSKRELMDLGLRLGADVVPVFGCSAMGEGVSEDLQAITLPEOWFVILKPDCHVNT 181
 122 WQGLSDMDLGLRLGADVVPVFGHAAPAEVGSGKLTPEAPAPVTVVLPQVSVTAEIFSDPL 181
 182 GSFSAENVLTNSAVVMSDFLAGNNDCEVVCCKLTPPVDAIDALLCYAARLTGACVAFQFCNK 241
 182 PVIFKDELPRNTKRSIFETLTKCFPSNDCEVLAARKFREVDATLSWLEAPAPRLTGTG 241
 242 ACVPAFCNKEDASALBGLKDRML-VFLAKGNOSALYKLE 282
 242 ACVPAFCNKEDASALBGLKDRML-VFLAKGNOSALYKLE 282

Query Match 38.6%; Score 584.5; DB 12; Length 280;
 Best Local Similarity 45.3%; Pred. No. 7.8e-55;

QY 11 WPAPAKINIMLRITGRRPDGYHLLQTVFQMLDLCMLTFHPVDDGAVTLNPISGVPEOD 70
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 21079
 ; LENGTH: 280
 ; TYPE: PRT
 ; ORGANISM: Xenorhabdus nematophilus
 ; US-10-369-493-21079

RESULT 7
 US-10-369-493-21079

Best Local Similarity 46.0%; Pred. No. 14e-54;
Matches 126; Conservative 41; Mismatches 105; Indels 2; Gaps 2

QY 11 WPAKAKNTLWRTGRSPDGTHLLQTFYFQWJLUDJCDMLTFHPVADGRV-TPLRNDISGVPQ 69
D 3 WPAKAKNTLQIYGRADGTHLLQTFYFRLLMDGDIYHRLKRDGSIQRTGASLPEVAD 62
QY 70 DDLTVRPAANTLKSHTCGVCRCYCIDIEKNLPMGGSLAGGSSDAATTLVLVNLWGLGSR 129
D 63 DDLWVRPAARLQIHAGTALGAEIRVYDKRIYAGGGFGGSSDAATVVALNALWGLPLPYD 122
QY 130 ELMDLGRLGADVPEVFYEGCSAMGEGVEDLOAITLPEQFVYIKPKDCYHTGRTPSAEN 189
D 123 TLMELGRLGADVPEVFYEGSNMAGVGELTTLTSLPQAAVYVDPGIVHVPVLPQOSE 182
QY 190 LTRNSAVVTMSDFLAGDNKDCSEVYCKLYRPVKDAI-DALLCYAEARLTGTGACVPAQF 248
D 183 LTRPAAPAKIADFASSGILDNAFEFVLRREPAIEAVFQLSRISGTPRLTGGSGCFVEF 242
QY 249 CNKEDAESALEGLKDRMLVFLAINGLQNSALYKXL 282
D 243 ATRPAABQMAHLPGNLRANVVEGAHSPILDL 276

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	38.3%;	Score 580.5;	DB 12;	Length 298;
Best Local Similarity	46.3%;	Pred. No. 2.3e-54;		
Matches 132;	Conservative 44;	Mismatches 106;	Indels 3;	Gaps 3

Qy	1	MDYAAAGGGEWSPAPAKINLMLRITGSRPDGYHLQTFQFQMLICMLITPHVDDCRV-TL	59
Qy	1	MDYAAAGGGEWSPAPAKINLMLRITGSRPDGYHLQTFQFQMLICMLITPHVDDCRV-TL	59
Db	11	VDDGVGVG-SAMPAPAKINLFLQITGRVGVGYHELQTFRELLIMGSLTHLRVDEGQIHRI	69
Qy	60	RNPISGVPEDDDLTYKPAANLTKSHTCSBQVCCIDIKNLPFMGGSGGGSDPAATLVYLA	119
Qy	70	GSEVTVGVVEADDLVYRPAAYILTKATVYHLGADIFVEKRIPIVGGSGGGSDPATLVYLA	129
Db	120	RWAGLQSKREMLDGLRLTGADYPRVPCGCSMGREVSVDLQATLPBQMFIIIPDCHV	179
Qy	130	ALMHTRLDVAVYLAALGRLGADYPRVHGGNAAABEGVCLTPMLIRGAAVILMDPGCV	189
Db	180	NGEITISAEKRLFNNSAVVTMSDFLAADNFNDCSE-VYCKLYRPYKDAIDALICVARELT	238
Qy	190	PIRELFLDPLRLFDASPAITIGDIFAGTAFGNAEPFLRRRSASVAGALCLVSEVGFARVY	249
Qy	239	GTGACVPAQFCNKEDAESALEGDKRWLVFLAKGINQSALYKKLE	283
Qy	250	GSQSGGFEVETSTDEAEKALERLPVGLCAWVADGASRSPDLVLVX	294

RESULT 12
US-10-369-493-9184

```

? Sequence 9184, Application US/10369493
? Publication No. US20030233675A1
? GENERAL INFORMATION:
? APPLICANT: Cao, Yongwei
? APPLICANT: Hinkle, Gregory J.
? APPLICANT: Slater, Steven C.
? APPLICANT: Goldman, Barry S.
? APPLICANT: Chen, Xiandeng
? TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
? TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
? FILE REFERENCE: 38-10 (52052) B
? CURRENT APPLICATION NUMBER: US/10/369,493
? CURRENT FILING DATE: 2003-02-28
? PRIOR APPLICATION NUMBER: US 60/360,039
? PRIOR FILING DATE: 2002-02-21
? NUMBER OF SEQ ID NOS: 47374
? SEQ ID NO 9184
? LENGTH: 270
? TYPE: PRT
? ORGANISM: Xylella fastidiosa
US-10-369-493-9184

```

Query Match	37.7%	Score 572;	DB 12;	Length 270;
Best Local Similarity	47.4%;	Pred. No. 1.7e-53;		
Matches 128;	Conservative 42;	Mismatches 98;	Indels 2;	Gaps 2

QY 11 WPAPALNTMLTITRRPFGYHLLQTVQMDDLCLMTFHPVDGVR-TLNTNPISGVEQ 69
QY 12 WPAPALNTMLTITRRPFGYHLLQTVQMDDLCLMTFHPVDGVR-TLNTNPISGVEQ 69
Db 1 WPAPALNTMLTITRRPFGYHLLQTVQMDDLCLMTFHPVDGVR-TLNTNPISGVEQ 60
QY 70 DDLTFRANMLKSHTCGVRGVCIDIEKNLPMNGGLGGSSDAATLVYLNRMGLSKR 129
Db 61 DDLVYRAAATLTKHANTNVYLGAIDIEPKRSIPVGGSGGSSDAATLVYLNRMGLSKR 120
QY 130 EELMDGLRAGADVPTFVFGSCSAMGGVSEDIQATLTPQWVLTIKPDCHVNTGTFSAEN 189
Db 121 VLAALGRLRGADVPTFVFGSCSAMGGVSECTTPMILPGAAVLLDPRGVCAPTRFLFDPD 180
QY 190 LFRNSAVVTMSDFLAGDNRNDCSE-VYCKLYRPVKDAIDALLCYAARLTGTGA CVFAQF 248
Db 181 LFRDASPAITGDFIGAAAFGNAFEPTVLRRESAVAGALDYLSEVGFARVWGSGGCEVEF 240
QY 249 CKKEQASALLEGKXDRMLVPLAKINGQSAL 278
Db 241 STRDSABCKLERLPGYGLCAWADGASRSP 270

```

RESULT 13
US-10-369-493-9359
; Sequence 9359, Application US/10369493
; Publication No. US2003023675A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
;
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;
; TITLES OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
;
; PIDS REFERENCE: 36-10(52052)B
;
; CURRENT APPLICATION NUMBER: US/10/369,493
;
; CURRENT FILING DATE: 2003-02-28
;
; PRIOR APPLICATION NUMBER: US 60/360,039
;
; PRIOR FILING DATE: 2002-02-21
;
; NUMBER OF SEQ ID NOS: 47374
;
; SEQ ID NO 9359
;
; LENGTH: 270
;
; TYPE: PRT
;
; ORGANISM: Xylella fastidiosa
;
; US-10-369-493-9359

```

Query Match 37.7%; Score 572; DB 12; Length 270;
 Best Local Similarity 47.4%; Pred. No. 1.7e-53;
 Matches 128; Conservative 42; Mismatches 98; Indels 2; Gaps 2;

QY 11 WPAPAKNLMRLITGRPDGYHLLQTVFQMLDLCMDLTFHPVDDGRV-TLNRPIISGVEEQ 69
 DB 1 WPAPAKNLMFLQITGRVGVYHLLQTVFRLDMGDTIHLRVBDEQIHRIGSVTGVVEA 60
 QY 70 DDLTAAANLKSHTGCVRGVCIDIEKULPMGGGIGGSSSDAATLVYLANLMLGSLSKR 129
 DB 61 DDLVRAAALLHARNVYIGADIPEKPIFVGSGGSSDAATLVLANLMLHTRLDVA 120
 QY 130 ELMDLGLRLGADVPVFFVSCSANGVSEDLQAITLPEQMFYIKPDCHVNTGEIFSAEN 189
 DB 121 VLAAGLRLGADVPVFFVSCSANGVSEDLQAITLPEQMFYIKPDCHVNTGEIFSAEN 180
 QY 190 LTRNSAVVTMSDFLAGDNNDCEB-VCKLTPPVKDAIDALCYEARLGTGACVFAQF 248
 DB 181 LTRDASPATIGDFIAGTAFGNAPFEPVLRRESAVAGALDMLSRVGFATVSGSCFVEF 240
 QY 249 CNKEDASALEGLKDRMLVFLAKGLNOSAL 278
 DB 241 STRDEHECHLERLPYGLCANVADGASRSL 270

RESULT 14
 US-10-369-493-19675
 ; Sequence 19675, Application US/10369493
 ; Publication No. US20030233675A1

GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 PRIOR FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 19675
 LENGTH: 213
 TYPE: PRT
 ORGANISM: Nitrosomonas europaea
 US-10-369-493-19675

Query Match 36.7%; Score 556; DB 12; Length 213;
 Best Local Similarity 51.6%; Pred. No. 6.6e-52;
 Matches 110; Conservative 34; Mismatches 67; Indels 2; Gaps 2;

QY 11 WPAPAKNLMRLITGRPDGYHLLQTVFQMLDLCMDLTFHPVDDGRVTLNRPIISGVEEQ 70
 DB 1 WPAPAKNLMFLQITGRVGVYHLLQTVFRLDMGDTIHLRVBDEQIHRIGSVTGVVEA 60
 QY 71 DDLTAAANLKSHTGCVRGVCIDIEKULPMGGGIGGSSSDAATLVYLANLMLGSLSKR 129
 DB 61 DDLVRAAALLHARNVYIGADIPEKPIFVGSGGSSDAATLVLANLMLHTRLDVA 120
 QY 130 ELMDLGLRLGADVPVFFVSCSANGVSEDLQAITLPEQMFYIKPDCHVNTGEIFSAEN 189
 DB 121 VLAAGLRLGADVPVFFVSCSANGVSEDLQAITLPEQMFYIKPDCHVNTGEIFSAEN 180
 QY 190 LTRNSAVVTMSDFLAGDNNDCEB-VCKLTPPVKDAIDALCYEARLGTGACVFAQF 248
 DB 181 LTRDASPATIGDFIAGTAFGNAPFEPVLRRESAVAGALDMLSRVGFATVSGSCFVEF 240

RESULT 15
 US-10-369-493-8419
 ; Sequence 8419, Application US/10369493

Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 PRIOR FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 8419
 LENGTH: 233
 TYPE: PRT
 ORGANISM: Ralstonia metallidurans
 US-10-369-493-8419

Query Match 35.2%; Score 533; DB 12; Length 233;
 Best Local Similarity 49.1%; Pred. No. 2.4e-49;
 Matches 114; Conservative 35; Mismatches 77; Indels 6; Gaps 2;

QY 53 DGRVTLNRPIISGVEEQDITVRAANLKSHTGCVRGVCIDIEKULPMGGGIGGSSDAA 112
 DB 2 DDGIVRFTDIPGPADTDLVRAAARMAQASGVFCVDIAIDKILPMGGGIGGSSDAA 61
 QY 113 TTLVLANLMLGSLSKRELMDLGLRLGADVPVFFVSCSANGVSEDLQAITLPEQMFYI 172
 DB 62 TTLALNRLMGVNLPREELMRIGLISGADVPPVFGNAFAEGVGBELTPVALPDSFFV 121
 QY 173 IKPDCHVNTGEIFSAENLTNSAVVTMSDFLAGDN-----RNDCEVCKLYRSPVKDAID 227
 DB 122 IHPQGHVPTAITSDELTNSPISIVADPACCTNKAFAFRNDLETATAKFGVAAALE 181
 QY 228 ALICYA-EARLTGACVFAQFCKEDASALSGLDRLVFLAKGLNOSAL 278
 DB 182 WLKNSPHAMWTGSGACVFAQFDEQVQRMERLPSEWDGRCVKSLSHHP 233

Search completed: January 29, 2004, 16:21:17
 Job time : 20.4547 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 8.38564 Seconds
(without alignments)
3268.453 Million cell updates/sec

Title: US-09-941-947a-12

Perfect score: 1516

Sequence: 1 MDYAGNGERMPAPAKNLNM.....LVPLAKINGSLYKLEEG 285

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: PIR_76:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	668	44.1	282	2	F83063 isopenylenyl monoph
2	652.5	43.0	299	2	AH0245 probable 4-diphosph
3	629	41.5	283	2	S27732 conserved hypotbet
4	629	41.5	283	2	AE0720 isopenylenyl monoph
5	624.5	41.2	281	2	A81875 hypothetical prote
6	616	40.6	283	2	A90843 ychb protein (simi
7	616	40.6	283	2	B47706 kinase, GMP famli
8	616	40.6	295	2	B81109 conserved hypotbet
9	615	40.6	281	2	B81149 hypothetical prote
10	609	40.2	283	2	P85700 conserved hypotbet
11	606.5	40.0	321	2	A64171 conserved hypotbet
12	580.5	38.3	298	2	P82532 isopenylenyl monoph
13	551	36.3	221	2	S49374 conserved hypotbet
14	493.5	33.6	294	2	G84949 hypothetical prote
15	327.5	21.6	299	2	A82654 hypothetical prote
16	327.5	21.6	299	2	F97435 conserved hypotbet
17	324	21.3	289	2	S66075 conserved hypotbet
18	323	21.3	291	2	AF1461 B. subtilis YabH p
19	323.5	21.3	293	2	AC3444 4-diphosphocytidyl
20	320	21.1	299	2	AG1098 B. subtilis YabH p
21	317	20.9	287	2	E83657 hypothetical prote
22	303	20.0	280	2	H89815 conserved hypotbet
23	301.5	19.9	282	2	A97257 isopenylenyl monoph
24	277	18.3	401	2	T07419 ripening-associate
25	275	18.1	383	2	T02642 probable ripening-4-diphosphocytidyl
26	272	17.9	317	2	AC2209 hypothetical prote
27	270.5	17.8	315	2	S74513 hypothetical prote
28	267	17.6	306	2	F70603 conserved hypotbet
29	260	17.2	299	2	E75254 conserved hypotbet

30	256.5	16.9	283	2	G81732 kinase, GMP famli
31	255.5	16.9	291	2	B71333 conserved hypotbet
32	254	16.8	311	2	B86939 probable isopenylen
33	250.5	16.5	268	2	A70379 conserved hypotbet
34	239	15.8	295	2	A87415 kinase, GMP famli
35	236	15.6	288	2	B71470 probable kinase -
36	233.5	15.4	271	2	H22261 conserved hypotbet
37	178	11.7	141	2	H86609 kinase (imported)
38	167	11.0	255	2	C81314 hypothetical prote
39	145.5	9.6	274	2	F71820 hypothetical prote
40	134	8.8	265	2	D82871 conserved hypotbet
41	132.5	8.7	322	2	F82583 conserved hypotbet
42	132	8.7	268	2	C64700 conserved hypotbet
43	125	8.2	314	2	F84952 homoserine kinase
44	119	7.8	309	2	F84952 homoserine kinase
45	116	7.7	990	2	JC7878 fucokinase (EC 2.7

ALIGNMENTS

RESULT 1

F83063 isopenylenyl monophosphate kinase PA4669 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: F83063

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Bri

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; Lim,

.; Lory, S.; Olson, M.V.

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathoc

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: F83063

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-282 <STO>

A:Cross-references: GB:AE004880; GB:AE004091; NID:G9950932; PIDN:AA08056.1; GSPDB:GN001:

A:Experimental source: strain PA01

C:Genetics:

A:Gene: ipk; PA4669

C:Superfamily: conserved hypothetical protein HP1443

Query Match	44.1%	Score 668	DB 2	Length 282
Best Local Similarity	51.8%	Pred. No. 1.1e-52		
Matches 141	Conservative 36	Mismatches 95	Indels 0	Gaps 0
QY	12	PAPAKNLMLRIRGRPRDGHILQTPQMDLGDWLFTHFVDDGRVTLRNPISGVPDOD	71	
DB	8	PAPAKNLMLRIRGRPRDGHILQTPQMDLGDWLFTHFVDDGRVTLRNPISGVPDOD	67	
QY	72	LTPAANILKSHGCVRGVCIDIEKNLPMGGGLGGSSDAATTLVNLRLMGLSLREL	131	
DB	68	LTPAANILKSHGCVRGVCIDIEKNLPMGGGLGGSSDAATTLVNLRLMGLSLREL	127	
QY	132	MDGLRLGAVPPVPPVPGCSAMGCVSEDLQATLPBQMFITIPDCHVNGEIPSAENLT	191	
DB	128	MDGLRLGAVPPVPPVPGCSAMGCVSEDLQATLPBQMFITIPDCHVNGEIPSAENLT	187	
QY	192	RNSAVVTMSDFLAGDNNDSEVVCCKLYRPVKAIDALLCYAARLTGTGACVPAQPCNK	251	
DB	188	RDSALIKVTRVLSGDSNDQPFVBERRYREVRNALILNKFASRLTGTGACVPAQPCNK	247	
QY	252	EDASALIEGLKDRMLVPLAKINGSLYKLE	283	
DB	248	AEADKVSALLPDLQRFVAAKSNISMLHRL	279	

RESULT 2

AH0245 probable 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.-) [imported] - Yers

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001

C/Accession: AH0245
R/Parikhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A/Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A/Reference number: AB0001; MUID:21470413; EUID:1156360
A/Accession: AH0245
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-299 <KUR>
A/Cross-references: GB:AL590842; PIDD:CA030828.1; PID:G15380029; GSPDB:GN00175
C/Genetics:
A/Gene: *ipk*
C/Superfamily: conserved hypothetical protein HP1443
C/Keywords: phosphotransferase

Query Match	43.0%;	Score 652.5;	DB 2;	Length 299;
Best Local Similarity	49.8%;	Pred. No. 3e-51;		
Matches 137;	Conservative 42;	Mismatches 91;	Indels 5;	Gaps 3

Qy	10	RNPAPAKNLMIRIGSRPDGYLLQVYFQKLDLMDLTSPVDDGVTLNPIISGVPEQ	69
Db	14	KNPSPAKNLFYITIGGRADYHQYQLFQGLTGVDDLTLEPRDDNQIRLLPTIAYENB	73
Qy	70	DDLTVEPANKLKSHTG--CVRGVCTIDENKLPWGGLGGSSSDAATTLVNLRLMGDL	126
Db	74	QNLIVFAAKMLQKHGNTFVRPGADISIDKCLPMGGGIGGSSNAAVTLVNLRLMQCL	133
Qy	127	SKRELMDLGLTGADVPVFPFGCSAMGSGVEDLQATLEPGWVITKPDCHVNTGEIFS	186
Db	134	TBEQLADLGLTGADVPVFPVFGHAFNFGIEENKQPAPEPEKMTLVHPEQNIPTIIFS	193
Qy	187	AENLTENSAVVTMSDFLAGNRENDSEVVCYLVRPYDAIDALLCAEAELITGTGACVFA	246
Db	194	DEELKNTPIRPLAALLSTPYANDCEPIARFRPREVEQALSWLLEYPASRLTGTGACVFA	253
Qy	247	QPCNKEDASAEGLKDRNL--VFLAKIGINOSALR	280
Db	254	EDDTSSAQVL-SLAPENLGFVARGNVSPLER	287

RESULT 3
S27732
conserved hypothetical protein 2 - *Salmonella typhimurium*
C:/Species: *Salmonella typhimurium*
C:/Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jan-2000
C:/Accession: S27732
R:/Post, D. A.; Hore-Jensen, B.; Switzer, R. L.
submitted to the EMBL Data Library, October 1991
A:/Description: The prs gene of *E. coli* and *S. typhimurium* is transcribed from two
A:/Reference number: S27568
A:/Accession: S27732
A:/Molecule type: DNA
A:/Residues: 1-283 <POS>
A:/Cross-references: EMBL:M77236; NID:g154280; PIDD:AAA27195.1; PID:g154282
C:/Superfamily: conserved hypothetical protein HP1443

Query Match	41.5%;	Score 629;	DB 2;	Length 283;
Best Local Similarity	50.2%;	Pred. No. 3.7e-49;		
Matches 140;	Conservative 35;	Mismatches 96;	Indels 8;	Gaps 4

[illegible][illegible]

RESULT 4

ALBU 120

C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
isopentenyl monophosphate kinase (EC 2.7.1.1) [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhi

A-Note: This species has also been called *Salmonella typhi*.
 C-Date: 09-Nov-2001 #sequence
 C-Date: 09-Nov-2001 #text
 C-Date: 18-Nov-2005 #change

R.Packhill, J.: Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, R.
CJAccession: A60720
R.Packhill, J.: Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, R.

S. J. Moule, S. J. O'Gaora, P. Nature 413, 848-852, 2001

A; Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A; Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar

A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AE0720
A;Status: preliminary

A:Residues: 1-283 <PAR>
A:Molecule type: DNA
A:status: preliminary

ACross-references: GB:AL513382; PIDM:CAD02134.1; PID:g16502968; GSPDB:GN00176
CGenetics:

A;Gene: 1pk
C;Superfamily: conserved hypothetical protein HP1443

C; Keywords: phosphotransferase

Query match	41.56	score 629	DB 27	length 263
Best Local Similarity	50.28	Pred. No. 3.7e-49		
Matches 140: Conservative	15	Mismatches	96	Indels 8
				Gaps 4

received	of	transmitted	of	more	of	exp
11	WPAEAKLNMRLRTGRRPDGYHLLQTVFOMLDCMLTTEHEVDDGRVTLRNEISGVEED	70				

Db 5 WPSAKNLFLYITGQRADSYHTQLFQGLDGYDTLIERHRDGEHLHLLFTVTGVENED 64

71 DLTVAANL!---KSHTCV---RGVCIDIEKNLEPMGGGLGGSSDAATTLVNLRIWGL 124

Db 65 NLVPAARLIMKVASSEGRIPAGGADISLEKRLPMGGGLCGSSNAAITVVALNHLMQC 124

125 GLEKREIMDLGLRLGADVPVFFGSGAMGCVSEDLQAITLPEQWVILKPDCHVNTGEI 164

Db 125 GSDIDELATLGLITGADVPFVRGHAAPAEAGVGELITPVNPPEKMYLVAHPGVSIPPIVI 184

185 PS:AEHLTENSAAVVTMSDFLAGDNNDSCSEVVCILYRPVKALIDALCYAEARLTGTGACV 244

D6 185 FKDPQLPNTPKRSIDVTLKCFBNDCEV LARKRFRBVDALSLWLE YAPSRHLGIGACV 244

245 FAFD CRRDABSRADSGKURWU - F LKRGKNGSRKLI RNU
262

[illegible]

RESULT 5
A81875

hypocnetical protein NMA1092 (imported) - Neisseria meningitidis (strain Z491 serogroup C)Species: Neisseria meningitidis
C Date of Nov 2000Neisseria meningitidis 03-Feb-2001

C:Accession: A61875
R:Parthill, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morello
C:Update: 03-May-2000 #sequence_revision 03-May-2000 #rev_change 02-Nov-2001

Hotroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Murgall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* 22491.
A:Reference number: A81775; PMID:20222556; PMID:10761919

A;Accession: A61875
A;Status: preliminary

A: molecule type: UNA

A:Residues: 1-281 <PAR>
A:Cross-references: GB:AI162755; GB:AI157959; NID:G7379742; PID:N:G7379757
A:Experimental source: serogroup A, strain Z491
C:Genetics:
A:Gene: NMA1092
A:Superfamily: conserved hypothetical protein HP1443

Query Match	41.2%;	Score 624.5;	DB 2;	Length 281;
Best Local Similarity	46.7%;	Pred. No. 9.3e-49;		
Matches 128;	Conservative 51;	Mismatches 88;	Indels 7;	Gaps 2

QY 9 ERMAPAPALNMLEITGRPPDGSHLQVFMELCDMLTFHPVDDGVLTANFIISGVPE 68
Dp 8 QAPAPAPALNMLDRLTRGRREDGYHNIBSIFCLIDLDQTVYLKPPDDGKILIHNEVDGMPQ 67
QY 69 QDDLTVRANMLKSHGTGVRGVCIDIEKNLPMGGSLGGSSDDAATTVLTVNRLMGLGSLK 128
Dp 68 EADLSYRPAASLLOKXARTPTGVEIWEIDKKIPLTGAGLGGSSDDAATTVLTVNRMQCGLTQ 127
QY 129 RELMDLGLRLGADVPVFWFGCSAMGEGVSEBDLOAITLPEQWFIIRKPKCHVNTGEIFSAE 188
Dp 128 RQLDISGALLGADVPFPIFGKNLPARGIGDLDEMDIPKQWYIVYKPPVHSTAKIPHSE 187
QY 189 NLTNNSAVVTWSDPLAGDN---ENDCSEVVCXLYRPAKDAIDALLCYAEARLLTGTGACV 244
Dp 188 GLTNSSASSIMPTF---QNLQPFRRNDQAVVFXKEPEWTKAYSHLSRGPALMTGSGACV 244
QY 245 FAQCFNKEDASALEGLKDRMLVFLAKINGSAL 278
Dp 245 FTACODNSAVINYIROVSDLTENYLAEGLSGEP 278

RESULT 6
 A90843
 Ycha protein [similarity] - Escherichia coli (strain O157:H7, substrain R1MD 0509552)
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001
 C:Accession: A90843
 R:Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G.,
 Gasasara, N., Iyanaga, T., Kuhtara, S., Shiba, T., Hattori, M., Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genoe
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: A90843
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-283 <HAV>
 A:Cross-references: CB:BA000007; P1DN:BA35136.1; P1D:G1361178; G55DB:GN00154
 A:Experimental source: Strain O157:H7, substrain R1MD 050952
 C:Genetics:
 A:Gene: ECE1713
 ;Superfamily: conserved hypothetical protein H21443

Query Match	40.6%;	Score 616;	DB 2;	Length 283;
Best Local Similarity	47.9%;	Pred. No. 5.5e-48;		
Matches 135;	Conservative 39;	Mismatches 96;	Indels 12;	Gaps 4;

[illegible]

```
Oy      242 ACVFAQFCNKBDAESALEGLKDRWL-VFLAKGLNGSALYTKL 282
          ||||| : : | : ||| | : :: :
Db      242 ACVFAPFDTSSEARQYLEQAPE-WLNGFVAKGANLSPLHRAM 282
```

RESULT 7
 B47706
 Y:Chn: protein - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 19-Dec-1993 #sequence revision 18-Nov-1994 #ext_change 01-Mar-2002
 C:Accession: B47706; E64867; F01178; 527569
 R:Post, D.A.; Hove-Jensen, B.; Switzer, R.L.
 J. Gen. Microbiol. 139, 259-266, 1993
 A:Title: Characterization of the hemA-prs region of the Escherichia coli and Salmonella typhimurium chromosomes
 A:Reference number: A47706; PMID:93171869; PMID:7679718
 A:Accession: B47706
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-283 <POS>
 A:References: EMBL:M77237; NID:G147380; PIDN:AAA24434.1; PID:G147382
 A:Experimental source: Strain K-12
 A:Note: Sequence extracted from NCBI Backbone (NCBIN:125627, NCBIP:125631)
 R:Schmaier, H.; Renner, P.; Hogenauner, G.
 J. Bacteriol. 177, 4488-4500, 1995
 A:Title: Expression of genes kdsA and kdsB involved in 3-deoxy-D-manno-octulosonic acid biosynthesis in Escherichia coli K-12
 A:Reference number: I60364; PMID:95362678; PMID:7543480
 A:Accession: 183566

A:Gene: ychB
A:Map position: 26 min
C:Superfamily: conserved hypothetical protein HP1443

[illegible]

QY 182 GEIISAENLTNSAVVMSDFLAGDNNDCEVCKLYRPVKDAIDALCYAERLTGTG 241
 DB 182 PVIIFKPELPRMTPKRSITFLTKCEFSNCEVIARRRFEVAVLSWILEFAPSLTGTG 241
 QY 242 ACVFAQFCCKEDASLEGLKDRML-VFLAKGINOSALYKYL 282
 DB 242 ACVFAEFDTSEARQVLEQAPF-MINGFAKAGANLSPILHRAM 282

RESULT 8

kinase, GMP family VC2182 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
 E82109
 C/Species: Vibrio cholerae
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C/Accession: E82109
 R/RefSeq: J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Chardson, D.; Ermolova, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A/Reference number: A82035; MUID:20406833; PMID:10952301
 A/Accession: E82109
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-295 <HE>
 A/Cross-references: GB:AE004289; GB:AE003852; NID:99656725; PIDN:AAF95327.1; GSPDB:GN001
 A/Experimental source: serogroup O1; strain N16961; biotype El Tor
 C/Genetics:
 A/Gene: VC2182
 A/Map position: 1
 C/Superfamily: conserved hypothetical protein HP1443

Query Match 40.6%; Score 616; DB 2; Length 295;
 Best local similarity 47.1%; Pred. No. 5,8e-48;
 Matches 188; Conservative 46; Mismatches 98; Indels 0; Gaps 0;
 QY 11 WPAPAKLMLRTGRRPDGYHLQTVFQMLCDMLTFHPVDDGRVTLNPIISGVPEQ 70
 DB 13 WSPAPAKLMLRTGRRPDGYHLQTVFQMLCDMLTFHPVDDGRVTLNPIISGVPEQ 72
 QY 71 DLTVRANLLKSHGTGCVRGVCIIEKNLPMGGGLGGSSDAATTLVNLRLMGLASRE 130
 DB 73 NLTYYKAAVMAKNAAGSPPLADILQHLKVLPMGGIGGGSSNAATTLVNLRLMGLASRE 132
 QY 131 LMDLGRRLADVPVYFPGCSAMGEGVSEDLQAITLPEQFVYIKDCHVNTGEISAE 190
 DB 133 LAEIGLALADVPVYFPGCSAMGEGVSEDLQAITLPEQFVYIKDCHVNTGEISAE 192
 QY 191 TNSAVVMSDFLAGDNNDCEVCKLYRPVKDAIDALCYAERLTGTGACVAPFCN 250
 DB 193 MENTPRDLASLTTPYENDCEKIVRSILYPRVDKQSLMLQYAPSRILQTSQVAFSS 252
 QY 251 KEDASALEGLKDRMLVFLAKGINOSALYKYL 282
 DB 253 RDXAQAVFAQLSDNVLAFVAGGRNVSPLKXTL 284

RESULT 9

conserved hypothetical protein NMB0674 [imported] - Neisseria meningitidis (strain MCS8)
 B81149
 C/Species: Neisseria meningitidis
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C/Accession: B81149
 R/RefSeq: H.J.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masigiani, V.; Pizzi, M.
 A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MCS8.
 A/Reference number: A81000; MUID:20175755; PMID:10710307
 A/Accession: B81149
 A/Status: preliminary
 A/Molecule type: DNA

A/Residues: 1-281 <TE>
 A/Cross-references: GB:AE002439; GB:AE002098; NID:57226100; PIDN:AAF41285.1; PID:9722611
 A/Experimental source: serogroup B, strain MCS8
 C/Genetics:
 A/Gene: NMB0874
 C/Superfamily: conserved hypothetical protein HP1443

Query Match 40.6%; Score 615; DB 2; Length 281;
 Best local similarity 46.1%; Pred. No. 6.7e-48;
 Matches 130; Conservative 53; Mismatches 91; Indels 8; Gaps 3;
 QY 1 MDVAGWGERPAPAKLMLRTGRRPDGYHLQTVFQMLCDMLTFHPVDDGRVTLR 60
 DB 1 MNIADG-RQAPSAPAKLMLRTGRRPDGYHLQTVFQMLCDMLTFHPVDDGRVTLR 59
 QY 61 NPIGVPEQDILTVRANLLKSHGTGCVRGVCIIEKNLPMGGGLGGSSDAATTLVNLRL 120
 DB 60 NPVGMFQEVDSLRASLLQKVRNPDVAGVETLWCKLPTGAGLGGSSDAATTLVNLRL 119
 QY 121 LMGGLSKRELMDGLRLGADVVPVPGCSAMGEGVSEDLQAITLPEQFVYIKDCHV 180
 DB 120 WMQGLRQRLIDGALLGADVPEFFIKKNAFARGIGBLDEKMDIPQWYVYKRPVHS 179
 QY 181 TGEIISAENLTNSAVVMSDFLAGDNNDCEVCKLYRPVKDAIDALCYAER 236
 DB 180 TAKIFHESLITNSASSIMPTF--QNLQEPFRMDQAVFEFEPYKAVSELSRYGAL 236
 QY 237 LTGTGACVFAQFCCKEDASLEGLKDRMLVFLAKGINOSAL 278
 DB 237 MTSGGACVFTACQDRNSAVNYRQVSDLYEAYLAEGLSKHL 278

RESULT 10

hypothetical protein ychb [imported] - Escherichia coli (strain O157:H7, substrain EDL93)
 F85700
 C/Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: F85700
 R/RefSeq: N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: F85700
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-283 <STO>
 A/Cross-references: GB:AE005174; NID:912514921; PIDN:AAG56066.1; GSPDB:GN00145; UMG:219
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Gene: ychb
 C/Superfamily: conserved hypothetical protein HP1443

Query Match 40.2%; Score 609; DB 2; Length 283;
 Best local similarity 47.5%; Pred. No. 2.4e-47;
 Matches 134; Conservative 39; Mismatches 97; Indels 12; Gaps 4;
 QY 10 WPAPAKLMLRTGRRPDGYHLQTVFQMLCDMLTFHPVDDGRVTLNPIISGVPEQ 69
 DB 4 QWSPAPAKLMLRTGRRPDGYHLQTVFQMLCDMLTFHPVDDGRVTLNPIISGVPEQ 63
 QY 70 DLTVRANLLKSHGTGCVRGVCIIEKNLPMGGGLGGSSDAATTLVNLRL 121
 DB 64 DNLIVRAARLMLKTAADSGRLPTG--SGANISIDKRLPMKGGIGGGSSNAATTLVNLRL 121
 QY 122 WGLGLSKRELMDGLRLGADVVPVPGCSAMGEGVSEDLQAITLPEQFVYIKDCHV 181
 DB 122 WQGLSLMDLAEKGLTLGADVVPVPRGHAFAEGVGSLITPVDPEKMYLVAAHGVSIPT 181
 QY 182 GEIISAENLTNSAVVMSDFLAGDNNDCEVCKLYRPVKDAIDALCYAERLTGTG 241
 DB 182 PVIIFKPELPRMTPKRSITFLTKCEFSNCEVIARRRFEVAVLSWILEFAPSLTGTG 241

QY 192 RNSAVTMSDFLAGNRNDCSEVCKLTPPVKDA 225
 DB 188 RDSFPAIKVTVLEGDSEKNDQCPVERRRYPVRNA 221

RESULT 14

hypothetical protein ychB [imported] - Buchnera sp. (strain Aps)

C/Species: Buchnera sp.

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C/Accession: G84949

R/Shigobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A/Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A

A/Reference number: A84930; NCBI:20445173; PMID:10993077

A/Accession: G84949

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-294 <STO>

A/Cross-references: GB:AP000398; GSPDB:GN00144

A/Experimental source: strain Aps

A/Genetics:

A/Gene: ychB; B0170

C/Superfamily: conserved hypothetical protein HP1443

Query Match 32.6%; Score 493.5; DB 2; Length 294;

Best Local Similarity 40.6%; Pred. No. 6,7e-37;

Matches 113; Conservative 48; Mismatches 108; Indels 9; Gaps 2;

QY 13 APAKINLMRTTGRPPDGHLLQTFQMDLDCMLTFHPVDGRVTLRNPISGVBDL 72

DB 15 SPAKINLFLVYTGRRKXGYNHIOQLTFQPDYGDQFKLIANKGNIELFTEKKI FNNVQNS 74

QY 73 TVRAANLTKSHGCVR-----GVCIDLEKNLPMGGGLGGSSDAATTLVLRNLMGL 124

DB 75 IITAAKLKXK-TALLQGLQNSYGAKITFLKKNIPMGSGGGSSDAATTLVLRNLMGL 133

QY 125 GLSKRELMDLGLRGAADVFPVFGCSANGSGVSDLOAITLPQWFIIPDCHVNTGEI 184

DB 134 QVTLKELSLGLRIGADVGFVWNTAVIEGIDILYPIVQKEKMYLVVYPCINISTRYM 193

QY 185 FSAENITRNSAVTMSDFLAGNRNDCSEVCKLTPPVKDAIDALLCYAARLTGACV 244

DB 194 FSSPFLMSNTAKSLQVLLKTPKDFENIARKQFPIKLLIRMLSSVAPSHMTGTGSCV 253

QY 245 PAQCNKEDASALBGLKDRWIFLAKGLNQSALYKEL 282

DB 254 FSEFDNKGSAQKIFSVLPKRVGFIKSNVSIPLHKT 291

RESULT 15

AB2654 hypothetical protein ipk [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C/Accession: AB2654

R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erege, G.; Gallet, W.; Grant, C.; Guenther, D.; Kutyavich, T.; Levy, R.; Li, M.; McClellan

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, B.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2577; NCBI:21608550; PMID:11743193

A/Accession: AB2654

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-299 <KUR>

A/Cross-references: GB:AB008688; PDB:1AL41648.1; PDB:1G1738989; GSPDB:GN00186

A/Experimental source: strain C58 (Dupont)

A/Genetics:

A/Gene: ipk

A/Map position: circular chromosome

C/Superfamily: conserved hypothetical protein HP1443

Query Match 21.6%; Score 327.5; DB 2; Length 299;

Best Local Similarity 35.6%; Pred. No. 6.8e-22;

Matches 99; Conservative 41; Mismatches 111; Indels 27; Gaps 10;

QY 13 APAKINLMRTTGRPPDGHLLQTFQMDLDCMLTFHPVDGRVTLRNPISGVBDL 68

DB 16 APAKINLALHTVTPQPAQVHLETLVFTTAAGDRIKDAATDSFISGPEGLILSGDS 75

QY 69 QDDLVTRANLTK--SHTG-CVRGVCIDLEKNLPMGGGLGGSSDAATTLVLRNLMGL 124

DB 76 GDNVTRARBDILRDLASTGQPARPVDIHLEKNLPVASGIGGSADAATLRGLRHWDA 135

QY 125 GLSKRELMDLGLRGAADVFPVFGCSANGSGVSDLOAIT-LPQWFIIPDCHVNTGEI 183

DB 136 AIAPEKKSIALKLGADVPMCLASRPLIARIGIDIEALITDLPILSWVLNPLKAVSTPE 195

QY 184 FSAENITRNSAVTMS-----DFLAGNRNDCSEVCKLTPPVKDAIDALLCYA 233

DB 196 IFRLLQNRVNPFLTPSTIGATTGMDFLA-QSRNDQPPAQLL-PEIGELTGLSEE 253

QY 234 EA---RLTGTA-C-VPAQCNKEDASALBGLKDRW 265

DB 254 GATLVFMGSGATCFGIFHSFDAQNAKNAETSLKRRPGW 291

Search completed: January 29, 2004, 15:55:31

Job time : 9.38564 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 5.36681 Seconds
(without alignments)
2497.314 Million cell updates/sec

Title: US-09-941-947a-12

Perfect score: 1516

Sequence: 1 MDYAGMGKRWPAKALNLM.....LVFLAKGLNQSLYKCLEOG 285

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	668	44.1	282	1	ISPE_PSEAE
2	665	43.9	289	1	ISPE_RALSO
3	652.5	43.0	299	1	ISPE_YERPE
4	644.5	42.5	284	1	ISPE_SHEON
5	642	42.3	290	1	ISPE_VIBVU
6	636	42.0	295	1	ISPE_VIBPA
7	629	41.5	283	1	ISPE_SALTI
8	629	41.5	283	1	ISPE_SALTI
9	624.5	41.2	281	1	ISPE_NEIMA
10	624	41.2	295	1	ISPE_PASNU
11	620	40.9	283	1	ISPE_ECOL6
12	616	40.6	283	1	ISPE_ECOL6
13	616	40.6	295	1	ISPE_VIBCH
14	615	40.6	281	1	ISPE_NEIMA
15	606.5	40.0	313	1	ISPE_HAETN
16	599.5	39.5	295	1	ISPE_XANAC
17	584	38.5	295	1	ISPE_XANCP
18	580.5	38.3	298	1	ISPE_XYLF4
19	513.5	33.9	304	1	ISPE_WIGER
20	493.5	32.6	306	1	ISPE_BUCAL
21	484.5	32.0	292	1	ISPE_BUCAL
22	354	23.4	296	1	ISPE_CHLUP
23	332	21.9	283	1	ISPE_STRA3
24	332	21.9	283	1	ISPE_STRA3
25	327.5	21.6	299	1	ISPE_AGR75
26	324	21.4	289	1	ISPE_BACSU
27	323	21.3	291	1	ISPE_LISIN
28	322.5	21.3	299	1	ISPE_BRUMU
29	321	21.2	282	1	ISPE_STRMU
30	320	21.1	283	1	ISPE_LISMO
31	317	20.9	287	1	ISPE_BACHD
32	314	20.7	288	1	ISPE_CLOPE
33	313.5	20.7	299	1	ISPE_BRUSU

34	308	20.3	287	1	ISPE_THERN	08765 thermanser
35	303	20.0	282	1	ISPE_STAM	093w8 staphylococ
36	302	19.9	282	1	ISPE_STAP	08cqu6 staphylococ
37	301.5	19.9	280	1	ISPE_CLOAB	097f51 clostridium
38	293	19.3	301	1	ISPE_RHIME	092rml rhizobium m
39	286.5	18.9	295	1	ISPE_RHIL0	0986c6 rhizobium l
40	280.5	18.5	311	1	ISPE_SYNEL	08d1j1 synecococc
41	280	18.5	280	1	ISPE_OCEH	08eu37 oceanobacti
42	277	18.3	401	1	ISPE_LYCES	093841 lycopersico
43	275	18.1	383	1	ISPE_ARATH	081014 arabidopsis
44	272	17.9	317	1	ISPE_ANASP	08ys61 anabaena sp
45	270.5	17.8	315	1	ISPE_SYNY3	097663 synecocyst

ALIGNMENTS

RESULT 1
ISPE_PSEAE STANDARD; PRT; 282 AA.

AC P42805; 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
GN ISPE_OP_P44669.
OC Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OX Pseudomonadaceae; Pseudomonas.
RN NCBI_TaxID=287;
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.V., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Labig K., Lam R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Kester J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
RT Nature 406:959-964(2000).
[2]
RC SEQUENCE OF 1-221 FROM N.A.
RX STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=95189718; PubMed=7883699;
RA Hungerer C., Troup B., Roemling U., Jahn D.;
RT "Regulation of the hema gene during 5-aminolevulinic acid formation
RT in Pseudomonas aeruginosa.";
RL Bacteriol. 177:1435-1443(1995).
CC -!- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
CC GROUP OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
CC D-erythritol.
CC -!- PATHWAY: Nontyrosine terpenoid biosynthesis pathway; fourth
CC step.
CC -!- SIMILARITY: BELONGS TO THE ISPE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, AE004880; AAC08056.1; --
CC EMBL, X82071; CAAS7570.1; --
CC F01, F83063; F83063.

```

DR PIR; S49374; S49374.
DR HAMAP; MF_00061; -, 1.
DR InterPro; IPR006204; GMP kinase.
DR InterPro; IPR004424; ISP.
DR Pfam; PF00288; GMP kinases; 1.
DR TIGRPFAMs; TIGR00154; ISP; 1.
DR Transferrase; Kinase; Isoprene biosynthesis; ATP-binding;
KM Complete proteome.
KT NP_BIND 95 105 ATP (POTENTIAL).
SQ SEQUENCE 282 AA; 30843 MW; 141EBD4CF9438D3 CRC64;

Query Match
Best Local Similarity 44.1%; Score 668; DB 1; Length 282;
Matches 141; Conservative 36; Mismatches 95; Indels 0; Gaps 0;

QY 12 PAPAKNTLMRTGRERPDGYHLLQTVFQMLDCMLTFHPVDGKVTLRNPISGVEPDD 71
DB 8 PAPAKNTLMRTGRERPDGYHLLQTVFQMLDCMLTFHPVDGKVTLRNPISGVEPDD 67
QY 72 LTVRAANLTKSHGTGVRGVCIDIEKNLPMGGGIGGSSDAATTLVNLRLMGLSKREL 131
DB 68 LTVRAANLTKSHGTGVRGVCIDIEKNLPMGGGIGGSSDAATTLVNLRLMGLSKREL 127
QY 132 MDGLRLGADVPVFGCSAMGEVSEDLQAITLPBQMFVIIKPDCHVNTGEIIFSAENLT 191
DB 128 AALGLRLGADVPVFGCSAMGEVSEDLQAITLPBQMFVIIKPDCHVNTGEIIFSAENLT 187
QY 192 RNSAVVTMSDF-----LAGDRNDCSEVCLYRPVKADAILLCYAEALITGACVFAFCNK 251
DB 188 RNSAVVTMSDF-----LAGDRNDCSEVCLYRPVKADAILLCYAEALITGACVFAFCNK 247
QY 252 EDASALBEGIKRMVLFLAKGLNQSLYKLE 283
DB 248 AEAQVSALEPDLQRFVAKGSIISLHRLKE 279

RESULT 2
ISPSE_RALSO STANDARD; PRT; 289 AA.
AC Q8Y2EO;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
GN (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangent S.,
RA Aylar M., Billaut A., Broctier P., Camus J.C., Cartolico L.,
RA Chandler M., Choine N., Claudel-Renaud C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schexer T.,
RA Siquier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).

-1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy
group of 4-diphosphocytidylyl-2C-methyl-D-erythritol (By
similarity).
-1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-
methyl-D-erythritol.
-1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway, fourth
step.
-1- SIMILARITY: BELONGS TO THE ISPSE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR ZMBL; AL646052; CAD13924.1; -.
DR HAMAP; MF_00061; -, 1.
DR InterPro; IPR006204; GMP kinase.
DR InterPro; IPR004424; ISP.
DR Pfam; PF00288; GMP kinases; 1.
DR TIGRPFAMs; TIGR00154; ISP; 1.
DR Transferrase; Kinase; Isoprene biosynthesis; ATP-binding;
KM Complete proteome.
KT NP_BIND 99 109 ATP (POTENTIAL).
SQ SEQUENCE 289 AA; 31257 MW; A046E75006587A1 CRC64;

Query Match
Best Local Similarity 43.9%; Score 665; DB 1; Length 289;
Matches 140; Conservative 34; Mismatches 93; Indels 6; Gaps 2;

QY 12 PAPAKNTLMRTGRERPDGYHLLQTVFQMLDCMLTFHPVDGKVTLRNPISGVEPDD 71
DB 12 PAPAKNTLMRTGRERPDGYHLLQTVFQMLDCMLTFHPVDGKVTLRNPISGVEPDD 71
QY 72 LTVRAANLTKSHGTGVRGVCIDIEKNLPMGGGIGGSSDAATTLVNLRLMGLSKREL 131
DB 72 LTVRAANLTKSHGTGVRGVCIDIEKNLPMGGGIGGSSDAATTLVNLRLMGLSKREL 131
QY 132 MDGLRLGADVPVFGCSAMGEVSEDLQAITLPBQMFVIIKPDCHVNTGEIIFSAENLT 191
DB 132 MDGLRLGADVPVFGCSAMGEVSEDLQAITLPBQMFVIIKPDCHVNTGEIIFSAENLT 191
QY 192 RNSAVVTMSDF-----LAGDRNDCSEVCLYRPVKADAILLCYAEALITGACVFAFCNK 245
DB 192 RNSAVVTMSDF-----LAGDRNDCSEVCLYRPVKADAILLCYAEALITGACVFAFCNK 251
QY 246 AQCCKEDASALBEGIKRMVLFLAKGLNQSL 278
DB 252 AFDRAEQAVADQVPSSEMGRCAGITHHPL 284

RESULT 3
ISPSE_YERPE STANDARD; PRT; 299 AA.
AC Q8Y2Y1;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
GN (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R.,
RA Prentice M.B., Sebatia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Baeham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltham T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford A.K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).

-1- SIMILARITY: BELONGS TO THE ISPSE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration

```

RA MEDLINE=22137663; PubMed=12142430;
RA Deng W., Burland V., Plunkett G., III, Boulton A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau E., Zhou S., Schwartz D.C.,
RA Retherston J.D., Lindler J.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Niles M.L., Watson J.S., Blattner F.R.,
RA Perry R.D.;
RA "Genome sequence of Yersinia pestis KIM";
RA J. Bacteriol. 184:4601-4611(2002).
CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy
CC group of 4-diphosphocytidyl-2C-methyl-D-erythritol (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-
CC methyl-D-erythritol.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
CC step.
CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ414151; CAC90828.1; -
CC EMBL: AE013832; AAM85852.1; -
CC PIR: AH0245; AH0245.
CC HAMAP: MF_00061; 1.
CC InterPro: IPR006204; GHMP_kinase.
CC InterPro: IPR004424; ISPE.
CC Pfam: PF00288; GHMP_kinases; 1.
CC TIGRPFAMs: TIGR00154; ispe; 1.
CC Transferrase; Kinase; Isoprene biosynthesis; ATP-binding;
CC Complete proteome.
CC NP_BIND 106 116 ATP (POTENTIAL).
CC SEQUENCE 299 AA; 32675 MW; E86651DA6BE0D658 CRC64;
SQ
Query Match 43.0%; Score 652.5; DB 1; Length 299;
Best Local Similarity 49.8%; Pred. No. 1e-51;
Matches 137; Conservative 42; Mismatches 91; Indels 5; Gaps 3;
QY 10 RNPAPKALMLRITRRPDGTHLQTVFOMLDCMLTFHVDGKRTLRNPISGVEQ 69
DB 14 KMPSPAKMLFYITGQADYHQLQTLFQFDYQGLTIEPRDQRLPLPIAGVNE 73
QY 70 DDLTVRAANLKSHTG---CVRGVCIDIERKLPMSGGIGSGSSDAATTVANLMLGGL 126
DB 74 QNLIVRAAKMLQKHPENTVPRGADISIKCLPMGGIGLGGSSNATVVALNLMGGL 133
QY 127 SKEIMDLGLRLGADVVFVFGSGAMGSGVSEDLQATLPEQFVILKPDCHVNTGEIFS 186
DB 134 TBEQLADLGLTIGADYVVFVFGGAAPABGIGELQPAEFERKMYLVIHGVNIPPIIFS 193
QY 187 AENLTNSAVVMTSDPLAGDNINCKSEVCKLRPVDAIDALLCYAARLITGACVFA 246
DB 194 DBELKNTVIRPAAALSTPPYANDEPIARKPRFEVQALSWLEYPARLITGACVFA 253
QY 247 QFCNKEDASALEGLKDRWL-VFLAKGLNQSALYK 280
DB 254 EFDTESSARQVL-SIAPEWLHGFAVGAVVSPFLHR 287
RESULT 4
ISPE SHEON STANDARD; PRT; 284 AA.
AC Q8BAR0;
DT 15-SEP-2003 (rel. 42, Created)
DT 15-SEP-2003 (rel. 42, Last sequence update)
DT 15-SEP-2003 (rel. 42, Last annotation update)
DE 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).

GN ISPE OR S03836.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=2229786; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Neilson K.E., Gaidos E.J., Neilson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Keyer T., Tsapin A., Scott J., Beaman M., Brinkac B., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impratti M., Lee K., Berry K., Lee C.,
RA Mueller J., Kouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldman I.V., Smith H.O., Venter J.C., Neilson K.E., Fraser C.M.,
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis";
RL Nat. Biotechnol. 20:1118-1123(2002).
CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy
CC group of 4-diphosphocytidyl-2C-methyl-D-erythritol (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-
CC methyl-D-erythritol.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
CC step.
CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE015816; AAN56813.1; -
CC TIGR: S03836; -
CC HAMAP: MF_00061; 1.
CC Pfam: PF00288; GHMP_kinases; 1.
CC TIGRPFAMs: TIGR00154; ispe; 1.
CC Transferrase; Kinase; Isoprene biosynthesis; ATP-binding;
CC Complete proteome.
CC NP_BIND 98 108 ATP (POTENTIAL).
CC SEQUENCE 284 AA; 30721 MW; F5A22B814242496 CRC64;
SQ
Query Match 42.5%; Score 644.5; DB 1; Length 284;
Best Local Similarity 49.1%; Pred. No. 5.1e-51;
Matches 135; Conservative 41; Mismatches 98; Indels 1; Gaps 1;
QY 11 WPAPAKNLMLRITRRPDGTHLQTVFOMLDCMLTFHVDGKRTLRNPISG-VEQ 69
DB 9 WPAPAKNLFLHNGRRSDGTHLQTLFQFVDCDQDFRTDPELILHSTMAVAVDS 68
QY 70 DDLTVRAANLKSHTG---CVRGVCIDIERKLPMSGGIGSGSSDAATTVANLMLGGLSKR 129
DB 69 DNLLIRAAKLSQATIGFGAGIWLDRLLPMGGIGLGGSSDAATTVANLMLMTQUSHD 128
QY 130 EIMDLGLRLGADVVFVFGSGAMGSGVSEDLQATLPEQFVILKPDCHVNTGEIFS 189
DB 129 ELAIGLGLGADVVFVFGFAAPAGVGERLQAVNPRLMVLVIAFDHVGTAIVAFQDPL 188
QY 190 LTRNSAVVMTSDPLAGDNINCKSEVCKLRPVDAIDALLCYAARLITGACVFA 249
DB 189 LPRNTPLKGLDPTLISQPMANDQGLVVSQPVAKAGLWLEYPARLITGACVFA 248
QY 250 NKEDASALEGLKDRWL-VFLAKGLNQSALYK 284
DB 249 SQQAALAAALALPSDMGQFAKKNISPLIYRLNR 283

RESULT 5

ISPE_VIBVU STANDARD; PRT; 290 AA.

ID ISPE_VIBVU

Q8DF6; 15-SEP-2003 (Rel. 42, Created)

DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)

DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).

ISPE OR VVI0256.

OS Vibrio vulnificus.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI_Taxid=672;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CMCP6;

RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.T., Moon Y.H., Jeong H.,

RA Choy H.E.;

RT "Complete genome sequence of Vibrio vulnificus CMCP6.";

RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy

CC group of 4-diphosphocytidyl-2C-methyl-D-erythritol (By

CC similarity).

CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-

CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-

CC methyl-D-erythritol.

CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth

CC step.

CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AEO16797; AAC08792.1; -

DR HAMAP; MF_00061; -; 1.

DR Pfam; PF00288; GHMP_kinases; 1.

DR TIGRPFam; TIGR00154; ispe; 1.

KM Transferase; Kinase; Isoprene biosynthesis; ATP-binding;

KM Complete proteome.

FT NP BIND 96 106 ATP (POTENTIAL).

SQ SEQUENCE 290 AA; 31657 MW; 27161EC31B1ED8D CRC64;

Query Match 42.3%; Score 642; DB 1; Length 290;

Best Local Similarity 48.9%; Pred. No. 8.9e-51;

Matches 133; Conservative 43; Mismatches 96; Indels 0; Gaps 0;

QY 11 WPAFAKNTLMRTIGRRPDGYHLLQTVFQMLDLCWLTFRPVDDGRTLNPISGVBEQD 70

DB 8 WPSPAKNTLFYINGRTDNGYHELFQTLFQPLDHDGDLTTANDSGHITLTPDIVDPAVQ 67

QY 71 DLTVRAANLKSRTGCGRGVCIIDIEKTLPMGGGLGSSGSAATTLVYANLMLGSKR 130

DB 68 NLWKAANLQKKTGCTLGANNHILANKLPMGGGIGSSGSAATTLVYANLMLGSKR 127

QY 131 LMDIGLRLGADVVPVFGSGAMGVSBDLQATLTPPEQFVILKPDCHVTGILFSAENL 190

DB 128 LMDIGLRLGADVVPVFGSGAMGVSBDLQATLTPPEQFVILKPDCHVTGILFSAENL 187

QY 191 LTRNSAVVTMSDPLAGNRNDCEVVCYLTPVYKDAIDALLCYARALITGACVYFQPCN 250

DB 188 LTRNPKRDEFTLTLNAPSVNDCCKIVRMILYPEVVKQSLWLLQYAPSRLLTGSCVFAEFS 247

QY 251 NKDAESALBGLKDRWLVEFLAKGNOSALYKGL 282

DB 248 RSESESTLQAQSDNVSFAVAGRNISPLKETT 279

RESULT 6

ISPE_VIBPA STANDARD; PRT; 295 AA.

ID ISPE_VIBPA

Q87RNT; 15-SEP-2003 (Rel. 42, Created)

DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)

DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).

ISPE OR VP0740.

OS Vibrio parahaemolyticus.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI_Taxid=670;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RIMD 221063 / Serotype O3:K6;

RX MEDLINE=22508454; Pubmed=12620739;

RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,

RA Iijima Y., Najiima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,

RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;

RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism

RT distinct from that of V. cholerae.";

RL Lancet 361:743-749(2003).

CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy

CC group of 4-diphosphocytidyl-2C-methyl-D-erythritol (By

CC similarity).

CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-

CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-

CC methyl-D-erythritol.

CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth

CC step.

CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AP005075; BACS9003.1; -

DR HAMAP; MF_00061; -; 1.

DR Transferase; Kinase; Isoprene biosynthesis; ATP-binding;

KM Complete proteome.

FT NP BIND 101 111 ATP (POTENTIAL).

SQ SEQUENCE 295 AA; 32497 MW; EF77FF8973EC076D CRC64;

Query Match 42.0%; Score 636; DB 1; Length 295;

Best Local Similarity 48.0%; Pred. No. 3.2e-50;

Matches 132; Conservative 46; Mismatches 97; Indels 0; Gaps 0;

QY 10 WPAFAKNTLMRTIGRRPDGYHLLQTVFQMLDLCWLTFRPVDDGRTLNPISGVBEQD 69

DB 12 WPSPAKNTLFYINGRTDNGYHELFQTLFQPLDHDGDLTTANDSGHITLTPDIVDPAVQ 71

QY 70 DLTVRAANLKSRTGCGRGVCIIDIEKTLPMGGGLGSSGSAATTLVYANLMLGSKR 129

DB 72 DLTVRAANLKSRTGCGRGVCIIDIEKTLPMGGGLGSSGSAATTLVYANLMLGSKR 131

QY 130 LMDIGLRLGADVVPVFGSGAMGVSBDLQATLTPPEQFVILKPDCHVTGILFSAENL 189

DB 132 LMDIGLRLGADVVPVFGSGAMGVSBDLQATLTPPEQFVILKPDCHVTGILFSAENL 191

QY 190 LTRNSAVVTMSDPLAGNRNDCEVVCYLTPVYKDAIDALLCYARALITGACVYFQPCN 249

DB 192 LTRNPKRDEFTLTLNAPSVNDCCKIVRMILYPEVVKQSLWLLQYAPSRLLTGSCVFAEFS 251

QY 250 NKDAESALBGLKDRWLVEFLAKGNOSALYKGL 284

DB 252 SKSESTLQAQSDNVSFAVAGRNISPLKETT 286

RESULT 7

ISPE_SALTY STANDARD; PRT; 283 AA.
 AC Q82699;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
 GN (4-cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase)
 DE ISPE OR IPK OR STY1905 OR IT1097.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 NCBI_TaxID=601;
 [1]

SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.W., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Hague A., Hien T.T., Holtroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RC "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 [2]

SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RC "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).

-1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy
 CC group of 4-diphosphocytidylyl-2C-methyl-D-erythritol (by
 CC similarity).
 CC CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-
 CC methyl-D-erythritol.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
 CC step.

-1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; AL627271; CAD02134.1; -;
 DR EMBL; A6016837; AA068760.1; -;
 DR HAMAP; MF_00061; -; 1.
 DR InterPro; IPR006204; GMP_kinase.
 DR InterPro; IPR004424; ISPE.
 DR Pfam; PF00288; GMP_kinases; 1.
 DR TIGRfam; TIGR00154; ispe; 1.
 KW Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
 KW Complete proteome.
 FT NP BIND 99 ATP (POTENTIAL).
 FT SEQUENCE 283 AA; 30870 MW; B3B9CB498A7F368 CRC64;

Query Match 41.5%; Score 629; DB 1; Length 283;
 Best Local Similarity 50.2%; Pred. No. 1,3e-49;
 Matches 140; Conservative 35; Mismatches 96; Indels 8; Gaps 4;

QY 11 WPAPKNTLMRTNRBPBGVHLLQTVFQMLDLCMLTFHPVDGRVTLNPDISGVPRD 70
 DB 5 WSPSPKNTLFLYITGQGRDGVHTLTQLTFQFLDYGTHTIIEPHDSEHLLFPVGVENED 64

QY 71 DLTVAAAILL---KSHGCV---RGVCIDIEKNLPMGGGLGGSSDAATTVLNLRLMGL 124
 DB 65 NLIVAAAILLKKVMSSEGRLLPAGSGADISIEKRLPMGGGLGGSSDAATTVLNLRLMGL 124

QY 125 GLSKRELMDLGRIGADYPPVPPVPGCSAMGEVSEDLQALTPBQWFLIKPDCHVNTGEI 184
 DB 125 GLSIDELATLGLTGADVPVPRGPAFAEGVGEITLTPVNPPEKMYLVAPHGVSTPTEVI 184

QY 185 PSAENVLTNSAVVTMSDPLAGNNRNDCEVCKYRPPKQIDALLCYAEKRLNTGACV 244
 DB 185 FKDPPLPNTKRSITLTKKEPNSDCVIAKRFREVDALLSWLLEAPSLTGTGACV 244

QY 245 PAQCNKEDAESALLEGKDRWL-VFLAKGLNQSLAYKXL 282
 DB 245 FAEPFTBSCARQVLEQAPR-WLNAFVAKVNLSPHREL 282

RESULT 8
 ISPE_SALTY STANDARD; PRT; 283 AA.
 AC P30753;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
 DE (4-cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase)
 GN ISPE OR IPK OR STY1779.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 NCBI_TaxID=602;
 [1]

SEQUENCE FROM N.A.
 RX MEDLINE=93171869; PubMed=7679718;
 RA Post D.A., Hove-Jensen B., Switzer R.L.;
 RT "Characterization of the hema-prs region of the *Escherichia coli* and
 RT *Salmonella typhimurium* chromosomes: identification of two open
 RT reading frames and implications for *prs* expression.";
 RL J. Gen. Microbiol. 139:259-266(1993).
 [2]

SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Forwellik S., All J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney B.,
 RA Ryan R., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RC "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 [3]

SEQUENCE OF 187-283 FROM N.A.
 RX MEDLINE=86257047; PubMed=2838463;
 RA Bower S.G., Hove-Jensen B., Switzer R.L.;
 RT "Structure of the gene encoding phosphoribosylpyrophosphate
 RT synthetase (*prsA*) in *Salmonella typhimurium*.";
 RL J. Bacteriol. 170:3243-3248(1988).
 CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
 CC GROUP OF 4-DIPHOSPHOCYTIIDYL-2C-METHYL-D-ERYTHRITOL (BY
 CC SIMILARITY).
 CC CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
 CC D-erythritol.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
 CC step.

-1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
 CC
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
 CC frameshift in position 268.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; W77236; AAA27195.1; -
DR EMBL; AE008779; AAL20694.1; -
DR EMBL; M19488; -; NOT_ANNOTATED_CDS.
DR PIR; S27732; S27732.
DR StyGene; SC100417; ispe.
DR HAMAP; MF_00061; -; 1.
DR InterPro; IPR006204; GHMP_kinase.
DR InterPro; IPR004424; ispe.
DR Pfam; PF00288; GHMP_kinases; 1.
DR TIGRfam; TIGR00154; ispe; 1.
DR Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
KM Complete proteome.
FT NP_BIND 99 109 ATP (POTENTIAL).
FT CONFLICT 252 252 S -> L (IN REF. 3).
SQ SEQUENCE 283 AA; 30883 MW; E33EAF956D12A5BD CRC64;

Query Match 41.5%; Score 629; DB 1; Length 283;
Best Local Similarity 50.2%; Pred. No. 1,3e-49;
Matches 140; Conservative 35; Mismatches 96; Indels 8; Gaps 4;

QY 11 WPAPAKLIMLRITGRPDGYHLLQTVFQMLDCWLTGHPVDDGAVTLRNPISGVPEOD 70
DB 5 WPSPAKLNFVLTGGRADGYHTLQTLFQFLDYGDTLHLEPHDDGHIHLLTFVNGVEMND 64
QY 71 DLTVRANLILKSHITGCV--RGVICIEKTLPMGGGLGGSSAAATTLVLRIMGL 124
DB 65 NLIVRARIIMKRVASSGRLLPAGSGADISIEKRLPMGGGLGGSSAAATTLVLRIMGL 124
QY 125 GUSKRELMDLGLRLGADVFFVFGCSAMGEVSEDLQATTLPEQMFVILKPDCHVNTGEI 184
DB 125 GUSIDELATLGLTGLADVFFVFGGHAFAEGVGEILITPVNPEKVLVHAPGVSIPTPVI 184
QY 185 FSEKENTRNSAVVTMSDFLAGDNNDCEVCKLTPVDAIDALLCYAEALITGACV 244
DB 185 FPDQPLPRTPKRSIDTLKCEPSNDEVIARRRFEVDAAISWLEVAPSRLTGACV 244
QY 245 PAQFCNKEDAESALGLKDRWL-VFLAKGLNOSALYKYL 282
DB 245 FAFOTESCARGVLEGAPE-WLNAFAKGVNLSPLHREL 282

RESULT 9
ISPE_NEIMA STANDARD; PRT; 281 AA.
ID ISPE_NEIMA
AC 09UK8;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
GN (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
NCBI_TaxID=56599;
RN SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D.T., Chillingworth T.,
RA Davis R.K., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
RA Jagers K., Leach S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,

RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506(2000).
CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
CC GROUP OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
CC D-erythritol.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
CC step.
CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AL162755; CAB84355.1; -
DR PIR; A61875; A61875.
DR HAMAP; MF_00061; -; 1.
DR InterPro; IPR006204; GHMP_kinase.
DR InterPro; IPR004424; ispe.
DR Pfam; PF00288; GHMP_kinases; 1.
DR TIGRfam; TIGR00154; ispe; 1.
DR Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
KM Complete proteome.
FT NP_BIND 98 108 ATP (POTENTIAL).
SQ SEQUENCE 281 AA; 31273 MW; E8A9AB81CCB40B80 CRC64;

Query Match 41.2%; Score 624.5; DB 1; Length 281;
Best Local Similarity 46.7%; Pred. No. 3.3e-49;
Matches 128; Conservative 51; Mismatches 88; Indels 7; Gaps 2;

QY 9 ERWPAPAKLIMLRITGRPDGYHLLQTVFQMLDCWLTGHPVDDGAVTLRNPISGVPE 68
DB 8 OAPAPAKLIMLRITGRPDGYHLLQTVFQMLDCWLTGHPVDDGAVTLRNPISGVPE 67
QY 69 QDDTLVRANLILKSHITGCV--RGVICIEKTLPMGGGLGGSSAAATTLVLRIMGL 128
DB 68 EADLSYRAASLLQYARTPTVATLWLDKIPYAGLGGSSDAATTLVLRIMGL 127
QY 129 RELMDLGLRLGADVFFVFGCSAMGEVSEDLQATTLPEQMFVILKPDCHVNTGEI 188
DB 128 RQIDSGALGLADVFFVFGGHAFAEGVGEILITPVNPEKVLVHAPGVSIPTPVI 187
QY 189 NITRNSAVVTMSDFLAGDN---ENDCEVCKLTPVDAIDALLCYAEALITGACV 244
DB 188 GITRNSASISIMPTF---QNLQPFRRNDQAVVFKKYPETWKAYSLSRGFALMTGSGACV 244
QY 245 PAQFCNKEDAESALGLKDRWL-VFLAKGLNOSAL 278
DB 245 FTACODRNSAINTYKQVSDIVYATILASLSHPL 278

RESULT 10
ISPE_PASMU STANDARD; PRT; 295 AA.
ID ISPE_PASMU
AC P57833;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
GN (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
NCBI_TaxID=747;


```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.,
RT "Complete genomic sequence of Pasteurella multocida pm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
CC GROUP OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL (BY
CC SIMILARITY).
CC CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
CC D-erythritol.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
CC step.
CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB006059; AAK02329.1; -.
CC DR HAMAP; MF 00061; -.
CC DR InterPro; IPR006204; GMP_kinase.
CC DR InterPro; IPR004424; ISPE.
CC DR Pfam; PF00288; GMP_kinases; 1.
CC DR TIGRFAMs; TIGR00154; ISPE; 1.
CC KM Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
CC Complete proteome.
CC NP BIND 108 118 ATP (POTENTIAL).
CC SQ SEQUENCE 295 AA; 33064 MW; 5FE9FEB7C9ABD55 CRC64;

Query Match 41.2%; Score 624; DB 1; Length 295;
Best Local Similarity 45.4%; Pred. No. 3.9e-49;
Matches 124; Conservative 49; Mismatches 100; Indels 0; Gaps 0;

QY 10 RMPAPAKLNLMTLRGRRPQGYHLQTVFQMDLCWMLTHFHPDQGVTLRNPISGVPCQ 69
DB 19 RFPAPAKLNLFLYINGRQGGYHEQLTFLQVFGDWLDEVEDNRCILTPELPSLKN 78
QY 70 DDLTVRAANLMLKSHTGCVRGVCIIEKNLPMGGGIGGSSDAATTVLNLKELSKR 129
DB 79 DNLVYRAAKLHQQKTNALGANLTLDKILPMGSGGGSSNATLVALNLTQULSTK 138
QY 130 ELMDLGLRLGADVVPVFGCSAMGEGVSEDLQATLTPQWFIIVKPDCHVNTGEIISAEN 189
DB 139 QLADGLMLGADVPIFVGHAAFAEGEKITCYCEPEKRYVVLKKNVSIATVFSDDP 198
QY 190 LTRSAVAVTMSDFLAGDNNDGSEVCKLYRPVDAIDALCYAARHGTGACVFAQPC 249
DB 199 LINTPQOSLEQLLNQKRYANDCEKYNLNPVEVEIIRLLQYAPSRSLTGACVFAERN 258
QY 250 NKEDASALEGLKDRMLVFLAKGINSALYKYL 282
DB 259 DEBSAQALFOTIPNYPFGVAAQGLNKSPLHNML 291

RESULT 11
ISPE_BCOL6 STANDARD; PRT; 283 AA.
AC Q8P104;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CKM)
DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
GN ISPE OR C166.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

```

```

OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:HI / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liu S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy
CC group of 4-diphosphocytidylyl-2C-methyl-D-erythritol (By
CC similarity).
CC CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-
CC methyl-D-erythritol.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
CC step.
CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB016759; AAN80131.1; -.
CC DR HAMAP; MF 00061; -.
CC DR Pfam; PF00288; GMP_kinases; 1.
CC DR TIGRFAMs; TIGR00154; ISPE; 1.
CC KM Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
CC Complete proteome.
CC NP BIND 99 109 ATP (POTENTIAL).
CC SQ SEQUENCE 283 AA; 30973 MW; E0327295B5EDDFC CRC64;

Query Match 40.9%; Score 620; DB 1; Length 283;
Best Local Similarity 47.9%; Pred. No. 8.4e-49;
Matches 135; Conservative 40; Mismatches 95; Indels 12; Gaps 4;

QY 10 RMPAPAKLNLMTLRGRRPQGYHLQTVFQMDLCWMLTHFHPDQGVTLRNPISGVPCQ 69
DB 4 QMPAPAKLNLFLYINGRQGGYHEQLTFLQVFGDWLDEVEDNRCILTPELPSLKN 63
QY 70 DDLTVRAANLMLKSHTGCVRGVCIIEKNLPMGGGIGGSSDAATTVLNLKELSKR 121
DB 64 DNLVYRAAKLHQQKTNALGANLTLDKILPMGSGGGSSNATLVALNLTQULSTK 121
QY 122 WGLGSKRLMDLGLRLGADVVPVFGCSAMGEGVSEDLQATLTPQWFIIVKPDCHVNT 181
DB 122 WCGGSLMDLAMLGTLGADVVPVFGHAAFAEGEKITCYCEPEKRYVVLKKNVSIATVFSDDP 181
QY 182 GEIPEANLITRNSAVVITMSDFLAGDNNDGSEVCKLYRPVDAIDALCYAARHGTGAC 241
DB 182 PVIFDPELPRNTPRISITLTKCEPSNDCEVIARRRERDAVAVSMLELVAFPSRLTG 241
QY 242 ACVFAEFTSEARQVLBOAPE-WLNGFVAAKGVNLSPIRRAM 282
DB 242 ACVFAEFTSEARQVLBOAPE-WLNGFVAAKGVNLSPIRRAM 282

RESULT 12
ISPE_BCOL1 STANDARD; PRT; 283 AA.
AC F24209;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CKM)

```

DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
 GN ISPE OR IPK OR B1208 OR Z1979 OR EC51713.
 OS *Escherichia coli*, and
 OS *Escherichia coli* 0157:H7.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; *Escherichia*.
 RX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=20040616; PubMed=10570138;
 RA Lange B.M., Croceau R.;
 RT "Isopentenyl diphosphate biosynthesis via a mevalonate-independent
 RT pathway: isopentenyl monophosphate kinase catalyzes the terminal
 RT enzymatic step.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13714-13719(1999).
 RN [2]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=K12 / W3110;
 RA Kuzuyama T., Takagi M., Kaneda K., Watanabe H., Dairi T., Seto H.;
 RT "Studies on the nonmevalonate pathway: Conversion of 4-(cytidine
 RT 5'-diphospho)-2-C-methyl-D-erythritol to its 2-phospho derivative by
 RT 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase.";
 RL Tetradon Lett. 41:2925-2928(2000).
 RN [3]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=20122571; PubMed=10655484;
 RA Luettgen H., Rohdich F., Herz S., Wungsinnawekul J., Hecht S.,
 RA Schuhr C.A., Fellemeier M., Sagner S., Zenk M.H., Bacher A.,
 RA Bismuth W.;
 RT "Biosynthesis of terpenoids: Ycbh protein of *Escherichia coli*
 RT phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2C-methyl-D-
 RT erythritol.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:1062-1067(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=93171869; PubMed=1679718;
 RA Post D.A., Hove-Jensen B., Switzer R.L.;
 RT "Characterization of the hemA-pvs region of the *Escherichia coli* and
 RT *Salmonella typhimurium* chromosomes: identification of two open
 RT reading frames and implications for pvs expression.";
 RL J. Gen. Microbiol. 139:259-266(1993).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93051347; PubMed=1427085;
 RA Ikemi M., Murakami K., Hashimoto M., Murooka Y.;
 RT "Cloning and characterization of genes involved in the biosynthesis
 RT of delta-aminolevulinic acid in *Escherichia coli*.";
 RL Gene 121:127-132(1992).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RA Remler P., Wolstschlaeger M., Strohmaier H.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Coliello-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1234-1238(1997).
 RN [8]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Ikemoto K., Inada H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Kimura S., Kitegawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,

RA Saito G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Pofstai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., DiSalvo E.T., Potamianus K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* 0157:H7.";
 RL Nature 409:529-533(2001).
 RN [10]
 RP SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT 0157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
 CC GROUP OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL. PHOSPHORYLATES
 CC ISOPENTENYL PHOSPHATE AT LOW RATES. ALSO ACTS ON ISOPENTENYL, AND,
 CC MUCH LESS EFFICIENTLY, DIMETHYLALLYL ALCOHOL. DIMETHYLALLYL
 CC MONOPHOSPHATE DOES NOT SERVE AS A SUBSTRATE.
 CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
 CC D-erythritol.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
 CC step.
 CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
 CC -1- CAUTION: Was originally (Ref.1) thought to be an isopentenyl
 CC monophosphate kinase.
 CC -1- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N- AND
 CC C-TERMINUS DUE TO FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AJ249325; CAB64963.1; -
 DR EMBL, AF179284; AAF13867.1; -
 DR EMBL, AF216300; AAF29530.1; -
 DR EMBL, AB037116; BA34247.1; -
 DR EMBL, W77237; AA24434.1; -
 DR EMBL, D10264; BAA01106.1; ALT_FRAME.
 DR EMBL, U18555; AAC43434.1; -
 DR EMBL, AE000219; AAC74292.1; -
 DR EMBL, D90756; BAA36066.1; -
 DR EMBL, AE005318; AAG56066.1; -
 DR EMBL, AP002556; BAB35136.1; -
 DR PIR, A30843; A30843.
 DR PIR, B47706; B47706.
 DR EcGene, EG11294; 1spB.
 DR HAMAP, MF_00061; -; 1.
 DR InterPro, IPR006204; GMP_kinase.
 DR InterPro, IPR004424; 1spB.
 DR Pfam, PF00288; GMP_kinases; 1.
 DR TIGRFAMs, TIGR00154; 1spB; 1.
 KW Transferase; Kinase; isoprene biosynthesis; ATP-binding;
 KW Complete proteome.

FT NP BIND 99 109 ATP (POTENTIAL).
 SQ SEQUENCE 283 AA; 30925 MW; 59A2921FA05D1B1 CRC64;
 Query Match 40.6%; Score 616; DB 1; Length 283;
 Best Local Similarity 47.9%; Pred. No. 1.9e-48;
 Matches 135; Conservative 39; Mismatches 96; Indels 12; Gaps 4;
 QY 10 RWPAPKLNLMITRRRPPGYHLQTVFQMLDLCMLTFHPVDDGRVTLRNISGVPEQD 69
 DB 4 QMSPPKLNLFYITGRADGYTLQTLFQFDYGTSTISLRLDDSDILRLPVEGVHS 63
 QY 70 DDLTVAANL-----KSHRGCVRGVCIDIEKNLPMGGGLGGSSDAATLVVNLRL 121
 DB 64 DNLIVRAARLMKTRADSGRLPTG--SGANISIDKRLPMGGGLGGSSNAATLVVNLRL 121
 QY 122 WGLGSLRELMDLGLGADVPVFGCSAMGSGVEDLQAITLPQCVFIIRKDCVNF 181
 DB 122 WCGGLSMDLELAEMGLTLGADVPVFGHAFAGVGELITVPDPKRYLVNHPVSIPT 181
 QY 182 GEIFSAENLTRNSAVVWMSDPLAGDNRNDCSEVVCCKLYRPVKAIDALLCYAEARLTGTG 241
 DB 182 PVIFKDELPFRNTPKSIETFLKCEFSNDCEVIAKRFREVDVLSMLLEYPASRLTGTG 241
 QY 242 ACVPAQCKNEDSALBGLKRWL-VFLAKGNQSAITYKL 282
 DB 242 ACVFAEDTSEARQVLEQAPR-WLNGFVAKGNLSPLHRAM 282
 RESULT 13
 ID ISPE_VIBCH STANDARD; PRT; 295 AA.
 AC Q9XQ23;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
 DS (4-cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
 GN ISPE OR VC2182.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrio.
 CX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RI TOR N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Emolaleva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Ueberlack T., Fleischmann R.D., Nieman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae.";
 RT cholerae.";
 RL Nature 406:477-483(2000).
 CC -!- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
 CC GROUP OF 4-DIPHOSPHOCYTIDYL-2-C-METHYL-D-ERYTHRITOL (BY
 CC SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
 CC D-erythritol.
 CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
 CC step.
 CC -!- SIMILARITY: BELONGS TO THE ISPE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AE004289; AAF95327.1; -
 DR PIR: B82109; B82109.
 DR TIGR: VC2182; -
 DR HAMAP: MF_00061; -; 1.
 DR InterPro: IPR006204; GMP Kinase.
 DR InterPro: IPR004424; ISPE.
 DR Pfam: PF00288; GMP Kinase; 1.
 DR TIGRFam: TIGR00154; ispe; 1.
 KM Transferrase, kinase; Isoprene biosynthesis; ATP-binding;
 KM Complete proteome.
 FT NP BIND 101 111
 SQ SEQUENCE 295 AA; 31973 MW; DF82A14981710531 CRC64;
 Query Match 40.6%; Score 616; DB 1; Length 295;
 Best Local Similarity 47.1%; Pred. No. 2e-48;
 Matches 128; Conservative 46; Mismatches 98; Indels 0; Gaps 0;
 QY 11 RWPAPKLNLMITRRRPPGYHLQTVFQMLDLCMLTFHPVDDGRVTLRNISGVPEQD 70
 DB 13 WSPAPKLNLFYITGRANGYHDLQTLFQFDHGDILTANSGNITLSPALADVALED 72
 QY 71 DLTVAANLKSHTCCVAGVCLDIKRNPMGGGLGGSSDAATLVNRLMGLSKRE 130
 DB 73 NLIVRAAALKNAAQSPLGADIQLKVLPNGGGLGGSSNAATLVNRYLQTLSDQ 132
 QY 131 LMDLGLRGADVPVFGCSAMGSGVEDLQAITLPQCVFIIRKDCVNFISAEVL 190
 DB 133 LAEIGLALGADVPVTRGRAPPAFAGVGELISVBERKYLIVRAVSLATDIFTHPL 192
 QY 191 TNSAVVWMSDPLAGDNRNDCSEVVCCKLYRPVKAIDALLCYAEARLTGTGACVPAQCN 250
 DB 193 MNNTFKRDLASLITLPYRNDCEKIVRSYLPEDKQISWLLQYAPRLTGTGSCVPAFSS 252
 QY 251 KEDASALEGLKRWLVFLAKGNQSAITYKL 282
 DB 253 KRDQAVPRAQSDNVLAVAGRNVSPLKTL 284
 RESULT 14
 ID ISPE_NEIMB STANDARD; PRT; 281 AA.
 AC Q9JZM4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
 DS (4-cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
 GN ISPE OR NMB0874.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 CC Neisseriaceae; Neisseria.
 CX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Nelson W.C., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Eisen J.A., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
 RA Cotton M.D., Ueberlack T.R., Khouri H., Qin H., Vamathevan J.,
 RA Gill J., Scarlato V., Maignant V., Pizzo M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 RT MC58.";
 RL Science 287:1809-1815(2000).
 CC -!- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
 CC GROUP OF 4-DIPHOSPHOCYTIDYL-2-C-METHYL-D-ERYTHRITOL (BY
 CC SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
 CC erythritol.

CC D-erythritol.
 CC -1- PATHWAY: Nomenclonate terpenoid biosynthesis pathway; fourth
 CC step.
 CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: A6002439; AAF41285.1; -
 CC PIR: B61149; B61149.
 CC TIGR: M30874; -
 CC HAMAP: MF_00061; -; 1.
 CC InterPro: IPR006204; GHMP_kinase.
 CC InterPro: IPR004424; ISPE.
 CC Pfam: PF00288; GHMP_kinases; 1.
 CC TIGRFAMs: TIGR00154; ispe; 1.
 CC Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
 CC Complete proteome.
 CC NP_BIND 98 108 ATP (POTENTIAL).
 CC FT SEQUENCE 281 AA; 31315 MW; FBFA5216CC2B7703 CRC64;
 CC
 CC Query Match 40.6%; Score 615; DB 1; Length 281;
 CC Best Local Similarity 46.1%; Pred. No. 2,4e-48;
 CC Matches 130; Conservative 53; Mismatches 91; Indels 8; Gaps 3;
 CC
 CC QY 1 MDYAAAGKGRKPAKAKLMLRTTGRBPDGTHLLQTVFQMLDLCWLTFFPVDGKVTIR 60
 CC DB 1 MNIADG-RQAFSAAPAKNLDLRTGRREDYHNIESIFCLIDIDQTVYVLEKPRDGIILH 59
 CC QY 61 NPISGVPEODDLTVRAANLIKSHNGCVRCVCIIEKNLPMWGGGLGGSSDAATTVLVAR 120
 CC DB 60 NPVDGMPQEVLDLTKASLILQKAKAPAGVEIWDKIPGLGAGLGGSSDAATTVLVAR 119
 CC QY 121 LMGGLSKRELMDLGLGADVVPFVFGCSAMGSEVSEDLQATLTFEQWVILKPDCHYN 180
 CC DB 120 WMCGGLTQROGLDISGALGADVPPFIFGKNAFARIGDRIDEMDIPQWTVIVKPVHVS 179
 CC QY 181 TGIFFSEKENTRNSAVVTVMSDFLAGDN---RNDGSEVYCKLYRPKDAIDALLCYAEAR 236
 CC DB 180 TAIIFTHESLTRNSASSIMPTF---QNLQPRNDQAVVKEPEWKAIVSELSRYGVAL 236
 CC QY 237 LMGTCACVFAQFCNKEDAESALGLKDRMLVFLAKGINOSAL 278
 CC DB 237 MTGSGACVFTACQDRNSAVTVVQVSLYENALYAEGLSKHPL 278
 CC
 CC RESULT 15
 CC ISPE_HAEIN STANDARD; PRT; 313 AA.
 CC ID ISPE_HAEIN STANDARD; PRT; 313 AA.
 CC AC P4527;
 CC DT 01-NOV-1995 (Rel. 32, Created)
 CC DT 20-NOV-1995 (Rel. 41, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
 CC DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
 CC ISPE OR H1608.
 CC OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 CC Pasteurellaceae; Haemophilus.
 CC NCBI_TaxID=727;
 CC RN
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=Rd / KW20 / ATCC 51907;
 CC MEDLINE=95350630; Pubmed=7542800;
 CC Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 CC Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 CC McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 CC Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Sadek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 CC RT Science 269:496-512(1995).
 CC CC
 CC CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
 CC GROUP OF 4-DIPHOSPHOCYTYDYL-2C-METHYL-D-ERYTHRITOL (BY
 CC SIMILARITY).
 CC CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
 CC D-erythritol.
 CC CC -1- PATHWAY: Nomenclonate terpenoid biosynthesis pathway; fourth
 CC step.
 CC CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC CC
 CC EMBL: U32834; AAC3252.1; -
 CC TIGR: H1608; -
 CC HAMAP: MF_00061; -; 1.
 CC InterPro: IPR006204; GHMP_kinase.
 CC InterPro: IPR004424; ISPE.
 CC Pfam: PF00288; GHMP_kinases; 1.
 CC TIGRFAMs: TIGR00154; ispe; 1.
 CC Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
 CC Complete proteome.
 CC NP_BIND 113 123 ATP (POTENTIAL).
 CC FT SEQUENCE 313 AA; 34657 MW; 7A84BAKAL196821B CRC64;
 CC
 CC Query Match 40.0%; Score 606.5; DB 1; Length 313;
 CC Best Local Similarity 47.1%; Pred. No. 1.6e-47;
 CC Matches 130; Conservative 45; Mismatches 100; Indels 1; Gaps 1;
 CC
 CC QY 10 RMPAPAKNLMRLTGRBPDGTHLLQTVFQMLDLCWLTFFPVDGKVTIRNPISGYPE 68
 CC DB 23 RPPSPAKNLMFLYINGKPPNGYHEQLTLPFLDFEDWDISIRBDDNOIVLPELPLMT 82
 CC QY 69 QDGLTVRAANLIKSHNGCVRCVCIIEKNLPMWGGGLGGSSDAATTVLVARLMGGLSK 128
 CC DB 83 ENNLTVRAAKLQBERANIGLANIHLDKILPMGGVGGSSDAATTVLVARLMGGLSK 142
 CC QY 129 RELMDGLRLGADVVPFVFGCSAMGSEVSEDLQATLTFEQWVILKPDCHYNVTSIFSAE 188
 CC DB 143 DELAKIGLTLGADVPIFVGHANAFAVGGEKITTCPEPAKRMVILKPDISTAVTFDDP 202
 CC QY 189 NITRNSAVVTVMSDFLAGDNNDGSEVYCKLYRPVDAIDALLCYAEARLTGTACVFAQF 248
 CC DB 203 NLPRTTPKRSIAQLLSBPYNDCEKVINHYSNEXALNMLQVAPALITGTGACVFAEF 262
 CC QY 249 CNKEDAESALGLKDRMLVFLAKGINOSALYKGEQ 284
 CC DB 263 DHEAEQAVFQKPEAFPGFVAKGINVSPRLHAKMQ 298
 CC
 CC Search completed: January 29, 2004, 15:50:40
 CC Job time : 5.36681 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 29, 2004, 15:44:14 : Search time 21.2436 Seconds
(without alignments)
3461.979 Million cell updates/sec

Title: US-09-941-947a-12
Perfect score: 1516
Sequence: 1 MEYAGWGBRMPAPAKNLIM.....LVEFLAKGINSALYKLESG 285

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	644.5	42.5	284	16	Q8EAP0
2	642	42.3	290	16	Q8DFP6
3	620	40.9	283	16	Q8F104
4	513.5	33.9	304	16	Q8D2K6
5	332	21.9	283	16	Q8E7K5
6	332	21.9	283	16	Q8E245
7	321	21.2	282	16	Q8E540
8	313.5	20.7	289	16	Q8E2D0
9	302	19.9	282	16	Q8CQ06
10	280.5	18.5	311	16	Q8DLJ1
11	280	18.5	280	16	Q8E0J3
12	274	18.1	401	10	Q8S2G0
13	223.5	14.7	315	16	Q8R0Z4
14	204.5	13.5	316	16	Q8E6I4
15	196	12.9	297	16	Q8E2M8
16	178	11.7	141	16	Q9J586

17	134	8.8	265	16	Q9PPN9
18	125.5	8.3	312	16	Q8E8T5
19	124.5	8.2	301	17	Q8PW39
20	119.5	7.9	286	17	Q8TU21
21	116	7.7	813	4	Q8N819
22	116	7.7	990	4	Q96MT9
23	114	7.5	306	16	Q8PLH7
24	114	7.5	370	10	Q8L7R2
25	111.5	7.4	288	16	Q8DZ12
26	110	7.3	160	10	Q8LKH4
27	110	7.3	370	10	Q9XE80
28	108.5	7.2	286	4	Q8TBE6
29	108	7.1	729	11	Q8C7W3
30	107.5	7.1	288	16	Q8E546
31	107.5	7.1	305	17	Q8TJ37
32	107.5	7.1	537	5	Q8I470
33	107	7.1	925	11	Q8C118
34	106	7.0	321	1	Q5J138
35	106	7.0	327	17	Q8PWJ3
36	105	6.9	288	2	Q9F6G9
37	104	6.9	318	16	Q8DEP3
38	102.5	6.8	335	2	Q8S011
39	102	6.7	288	2	Q9F6H0
40	102	6.7	288	2	Q9ET86
41	102	6.7	288	2	Q9ET82
42	102	6.7	288	2	Q9F6G7
43	102	6.7	288	2	Q9F6G0
44	102	6.7	288	2	Q9F6F4
45	102	6.7	288	2	Q9ET85

ALIGNMENTS

RESULT 1

Q8EAP0 PRELIMINARY; PRT; 284 AA.

AC Q8EAP0; 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE 4-diphosphocytidylyl-2C-methyl-D-erythritol kinase.
 GN 1SP8 OR S03836.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 RA Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Impraim N., Lee K., Berry K., Lee C.,
 RA Weller J., Kouti H., Gill J., Utterback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 Shewanella oneidensis."
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL; AB015816; AAN56813.1; -.
 DR TIGR; S03836; -.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 284 AA; 30721 MW; FES22B814242496 CRC64;

Query Match 42.5%; Score 644.5; DB 16; Length 284;

Best Local Similarity 49.1%; Pred. No. 1,5e-52;
 Matches 135; Conservative 41; Mismatches 98; Indels 1; Gaps 1;

QY 11 WPAKXNLMIRITGRRDGTHLLQTVQMDLDCDWLTFHVVDDGRVTLRNPISG-VPEQ 69

Db 9 WPAPAKLNLFLHNGRSDGYHELQTLFQVDCDDIDRVTDTPELIIHSTKSAVADS 68
 QY 70 DDLTVRAANLILKSHSGCVAGVCIDIEKNLPMGGGLGSSSDAATTLVNLRLMGLSKRE 129
 Db 69 DNLILRAAKSLQCATGFNGGAEIWLDRKLPMSGGLGSSSDAATTLVNLRLMNTQLSHD 128
 QY 130 ELMDLGLRLGADVPVVFPGCSANGGVSEDLQATLTPGQWFIKPDCHVNGEIFSAN 189
 Db 129 ELAIGIKLGLADIPVFIHGFAFGVGERLQAVNPAELMTYIADPAHVSTDAVHODPL 188
 QY 190 LITNSAVVTMSDFLAGDNNDCEVWCXLYPFKDAIDALLCYAEARLTGTGACVAPQPC 249
 Db 189 LPENTRLKGLDITLSSGFANDCCGLVSKYPOVAKLGLVLEAPSRMTGTACVAGERS 248
 QY 250 NKEDASALEGLKDRMLVFLAKLINGSALYKKEQ 284
 Db 249 SQOQALAAALAKLPDSMDQGFVAKMKNISPLIVLRN 283

RESULT 2

Q8DF6 PRELIMINARY; PRT; 290 AA.
 AC Q8DF6;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate synthase.
 GN V110256.
 OS *Vibrio vulnificus*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; *Vibrio*.
 OX NCBI_Taxid=672;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMCP6;
 RA Khee U.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
 RA Cho Y.H.E.;
 RT "Complete genome sequence of *Vibrio vulnificus* CMCP6."
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBS databases.
 DR EMBL; AB016797; AAO08792.1; -
 KW Complete proteome.
 SQ SEQUENCE 290 AA; 31657 MW; 27161EC131ED8D CRC64;

Query Match 42.3%; Score 642; DB 16; Length 290;
 Best Local Similarity 48.9%; Pred. No. 2,7e-52;
 Matches 133; Conservative 43; Mismatches 96; Indels 0; Gaps 0;

QY 11 WPAPAKLNLMLRTTGRPPDGYHLQTVFQMLDCMLTFHPVDDGRVTIANPISGVPEQ 70
 Db 8 WPSPAKLNLFYNGRTDNGYHELQTLFQLDHQDLTTANDSGHITLTPDIIVDLVBO 67
 QY 71 DLTVRANLILKSHSGCVAGVCIDIEKNLPMGGGLGSSSDAATTLVNLRLMGLSKRE 130
 Db 68 NLTWKANLILKSHSGCVAGVCIDIEKNLPMGGGLGSSSDAATTLVNLRLMGLSKRE 127
 QY 131 LMDLGLRLGADVPVVFPGCSANGGVSEDLQATLTPGQWFIKPDCHVNGEIFSAN 190
 Db 128 LADIGIKLGLADVPVFIHGFAFGVGERLQAVNPAELMTYIADPAHVSTDAVHODPL 187
 QY 191 LITNSAVVTMSDFLAGDNNDCEVWCXLYPFKDAIDALLCYAEARLTGTGACVAPQPC 249
 Db 188 LPENTRLKGLDITLSSGFANDCCGLVSKYPOVAKLGLVLEAPSRMTGTACVAGERS 247
 QY 251 NKEDASALEGLKDRMLVFLAKLINGSALYKKE 282
 Db 248 SQOQALAAALAKLPDSMDQGFVAKMKNISPLIVLRN 279
 RESULT 3
 Q8F104 PRELIMINARY; PRT; 283 AA.
 AC Q8F104;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148).
 GN ISPE OR C1666.
 OS *Escherichia coli* O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_Taxid=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plumkett G. III, Redford P., Roesch P.,
 RA Raeko D., Buckles E.L., Iden S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mobley G.F., Rose D.J., Zhou S., Schartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic *Escherichia coli*."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 DR EMBL; AB016759; AAN0131.1; -
 KW kinase; Transferase; Complete proteome.
 SQ SEQUENCE 283 AA; 30973 MW; E032729585EDDFC CRC64;

Query Match 40.9%; Score 620; DB 16; Length 283;
 Best Local Similarity 47.9%; Pred. No. 3.1e-50;
 Matches 135; Conservative 40; Mismatches 95; Indels 12; Gaps 4;

QY 10 WPAPAKLNLMLRTTGRPPDGYHLQTVFQMLDCMLTFHPVDDGRVTIANPISGVPEQ 69
 Db 4 QMPSPAKLNLFYNGRTDNGYHELQTLFQLDHQDLTTANDSGHITLTPDIIVDLVBO 63
 QY 70 DDLTVRAANLILKSHSGCVAGVCIDIEKNLPMGGGLGSSSDAATTLVNLRL 121
 Db 64 DNLILRAAKSLQCATGFNGGAEIWLDRKLPMSGGLGSSSDAATTLVNLRL 121
 QY 122 WGLSKRELMDLGLRLGADVPVVFPGCSANGGVSEDLQATLTPGQWFIKPDCHVNT 181
 Db 122 WQCGISMDLGLRLGADVPVFIHGFAFGVGERLQAVNPAELMTYIADPAHVSTDAVHODPL 181
 QY 182 GEIPSAEVLITNSAVVTMSDFLAGDNNDCEVWCXLYPFKDAIDALLCYAEARLTGTG 241
 Db 182 PVTFQDPELPNTGRPSIETLTKCESNDCEVIRARKRERDAVLSWLEAPSRMTGTG 241
 QY 242 ACVPAQCNKEDASALEGLDRMLVFLAKLINGSALYKKE 282
 Db 242 ACVPAQCNKEDASALEGLDRMLVFLAKLINGSALYKKE 282

RESULT 4
 Q8D2K6 PRELIMINARY; PRT; 304 AA.
 AC Q8D2K6;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE YchB protein.
 GN YCHB.
 OS *Mycobacterium tuberculosis* H37Rv.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Mycobacterium*.
 OX NCBI_Taxid=164609;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=22297718; PubMed=12219091;
 RA Akman I., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
 RA Aksoy S.;
 RT "Genome sequence of the endocellular obligate symbiont of tsetse
 fly, *Mycobacterium tuberculosis* H37Rv."
 RL Nat. Genet. 32:402-407 (2002).
 DR EMBL; AB063522; BAC24494.1; -
 KW Complete proteome.
 SQ SEQUENCE 304 AA; 34682 MW; FB746B5858189985 CRC64;

Query Match 33.9%; Score 513.5; DB 16; Length 304;
 Best Local Similarity 39.3%; Pred. No. 3.9e-40;
 Matches 110; Conservative 63; Mismatches 98; Indels 9; Gaps 2;

QY 11 WPAPAKINLMRTGRPDGTHLQTFQWLDLQWLTFFHVPDGRVTLNRPISGVPEOD 70
 Db 18 WEPAPAKINLFSLSTYGRKQGYHPIQTLFQFLANGDYILFPTTSDKTKLKHGRNEN 77

QY 71 DLTVRAANLI-----KSHTCGVRGVCIDIEKNLPMGGGLGGSSDAATTLVLRNLWG 123
 Db 78 NLITIRAKSKDKPMWKKKND--TPGVKIFIKKYIPIGGGLGGSSNMAATTLTALNENHK 135

QY 124 LGSKREIMDLGRIGADVPFVFGSAMGEGSEDIQATTLPEOMFVILKPDCHVNTGB 183
 Db 136 CKSLNTLADLQIGIDIVFYGKSAFAGIGELSLFQPEKRYLIVIPKISTKL 195

QY 184 IFSANLITRNSAVVMSDFLAGNRNDCSEVCKLYRVPKDAIDALLCYAARLTGTGAC 243
 Db 196 IFNKEILNKSYLSCVQYLKKPLKNDFBPMIRKCNFIIDNLINYSKSNFRLTGTGSC 255

QY 244 VFAQPCNKEDASALBGLKRWLVFLAKGLNOSALTKLE 283
 Db 256 IFSEFSECKAKKILYKLPKIKGVSKGTNISYLKELTK 295

RESULT 5
 OEB7K5 PRELIMINARY; PRT; 283 AA.

AC OEB7K5;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
 DE Hypothetical protein.
 GN GBS0149.
 OS Streptococcus agalactiae (serotype III).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=216495;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NEM316 / Serotype III;
 RX MEDLINE=22242508; PubMed=12354221;
 RA Glaeser P., Kusniox C., Buchrieser C., Chevallier F., Frangeul L.,
 RA Maeder T., Zouine M., Couve E., Lailouli E., Poyart C., Tilleu-Cuot P.,
 RA Kunst F.;
 RT "genome sequence of Streptococcus agalactiae, a pathogen causing
 RT invasive neonatal disease.";
 RT Mol. Microbiol. 45:1489-1513 (2002).
 DR EMBL; AL766844; CAD45794.1; --
 DR Sagalst; gbs0149; --
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 283 AA; 31075 MW; D195DA8A84CC0FE0 CRC64;

Query Match 21.9%; Score 332; DB 16; Length 283;
 Best Local Similarity 32.2%; Pred. No. 4.8e-23;
 Matches 85; Conservative 58; Mismatches 107; Indels 14; Gaps 5;

QY 13 APAKINLMRTGRPDGTHLQTFQWLDLQWLTFFHVPDGRVTLNRPISGVPEOD 72
 Db 7 APAKINLGLDIIKRCDDGYHELMIWISIDLNDYVITSELKEDCIVIDSDSKMPLNND 66

QY 73 TV-RAANLKSHTGVRGVCIDIEKNLPMGGGLGGSSDAATTLVLRNLWGGLSKREL 131
 Db 67 DFKKADITIKNGYINKGVHRLKSTIPVCAIGGSDTAATIRALNRLNMLQMDYDEM 126

QY 132 MDGLRLGADVPFVFGSAMGEGSEDIQATTLPEOMFVILKPDCHVNTGBIF--SA 187
 Db 127 VALGPKIGSVVPCGLGGGLVAGKEIVKPLPTLRPCWLVKPDGISTKSIFRDIDC 186

QY 188 ENLIR-----NSAVVMSDFLAGNRNDCSEVCKLYRVPKDAIDALLCYA---EARLT 238
 Db 187 KSISRVDDILKSAIISDYQLMVKWSGNSLEDITTKNPKVISTIKERLNSGADVAMT 246

QY 239 GTGACVFAQPCNKEDASALBGLK 262
 Db 247 GSGPTVFSMCTEKADRVFNSMK 270

RESULT 6
 OEB245 PRELIMINARY; PRT; 283 AA.

AC OEB245;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
 DE 4-diphosphocytidyl-2C-methyl-D-erythritol kinase.
 GN ISP OR SAG0153
 OS Streptococcus agalactiae (serotype V).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=216466;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2603 V/R / Serotype V;
 RX MEDLINE=22222988; PubMed=12200547;
 RA Tettelin H., Mastignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
 RA Madoff L.C., Wolf A.M., Beaman M.J., Brinkac L.M., Daugherty S.C.,
 RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
 RA Radune D., Fedorova N.B., Scanlan D., Kouri H., Mulligan S.,
 RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
 RA Jacobini E.T., Breton C., Galli G., Mariani M., Vegni F., Malone D.,
 RA Rinaldo D., Rappelli R., Telford J.L., Kasper D.L., Grand G.,
 RA Frazer C.M.;
 RT "Complete genome sequence and comparative genomic analysis of an
 RT emerging human pathogen, serotype V Streptococcus agalactiae.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
 DR EMBL; AB014198; AA09061.1; --
 DR TIGR; SAG0153; --
 KM Kinase; Complete proteome.
 SQ SEQUENCE 283 AA; 31047 MW; FC93843F69C2DC0D CRC64;

Query Match 21.9%; Score 332; DB 16; Length 283;
 Best Local Similarity 32.2%; Pred. No. 4.8e-23;
 Matches 85; Conservative 58; Mismatches 107; Indels 14; Gaps 5;

QY 13 APAKINLMRTGRPDGTHLQTFQWLDLQWLTFFHVPDGRVTLNRPISGVPEOD 72
 Db 7 APAKINLGLDIIKRCDDGYHELMIWISIDLNDYVITSELKEDCIVIDSDSKMPLNND 66

QY 73 TV-RAANLKSHTGVRGVCIDIEKNLPMGGGLGGSSDAATTLVLRNLWGGLSKREL 131
 Db 67 DFKKADITIKNGYINKGVHRLKSTIPVCAIGGSDTAATIRALNRLNMLQMDYDEM 126

QY 132 MDGLRLGADVPFVFGSAMGEGSEDIQATTLPEOMFVILKPDCHVNTGBIF--SA 187
 Db 127 VALGPKIGSVVPCGLGGGLVAGKEIVKPLPTLRPCWLVKPDGISTKSIFRDIDC 186

QY 188 ENLIR-----NSAVVMSDFLAGNRNDCSEVCKLYRVPKDAIDALLCYA---EARLT 238
 Db 187 KSISRVDDILKSAIISDYQLMVKWSGNSLEDITTKNPKVISTIKERLNSGADVAMT 246

RESULT 7
 OEB240 PRELIMINARY; PRT; 282 AA.

AC OEB240;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
 DE Putative isopentenyl monophosphate kinase.
 GN IPK OR SMU.1996.

OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 CX NCBI_TaxID=1309;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=UA159 / ATCC 700610 / Serotype C;
 RX MEDLINE=22295063; PubMed=12397186;
 RA Ajdic D., Mcshan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
 RA Li S., Zhu H., Najaf F., Lai H., White J., Roe B.A., Ferretti J.D.;
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
 pathogen."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
 DR EMBL; AE015023; AAN59600.1;
 KW Kinase; Complete proteome.
 SO SEQUENCE 282 AA; 30964 MW; 65FDJ629BFB84289 CRC64;
 Query Match 21.2%; Score 321; DB 16; Length 282;
 Best Local Similarity 32.5%; Pred. No. 5,2e-22;
 Matches 87; Conservative 57; Mismatches 102; Indels 22; Gaps 9;
 QY 13 APATLMLRTGRRPPGTHLQTFVOMLDCMLTFHPVDDGRVTLRNPISGV-EGD 70
 DB 7 APAKINFLDTLFRKDDGHEIEVMVTVLNDRLSPERKDNKIVDIEHNVFNDKN 66
 QY 71 DLTVRAANILKSHHTGCVRGVCIIDIEKNLPMGGGLGGSSDAATLVVLRMLGSLSKRE 130
 DB 67 DV-YKAAHLKRRRHISTGDKLSIQKIPICAGIGSSSDAAATLRLNCLMKNLSPKE 125
 QY 131 LMDIGRLGADVVFV-FGGS-AWGEVSEDLQATLPEQWFTIIRKDCVNTGEIFSAE 188
 DB 126 LIDPFIGSDVPYCIEAGCALISGKEIVPELATTL-STWVVLVKDFPISITTFKEI 184
 QY 189 NLTNSAV-----VTMSDF-----LAGDNRDCEVCKLYRPVKAIDALLCYAE- 234
 DB 185 DMAIISKVDLPALKEALIANYEALQPMGNSLEDITAKKPKIQRNM-IKGADI 242
 QY 235 ARLTGTACVFAQPCNKEDAESALEGLK 262
 DB 243 ALMTGSGTVFALCRTERKADRVVSMK 270
 RESULT 8
 Q8G2D0 PRELIMINARY; PRT; 299 AA.
 ID Q8G2D0
 AC Q8G2D0;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE 4-diphosphocytidylyl-2C-methyl-D-erythritol kinase.
 GN ISPE OR BR0394.
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 CX NCBI_TaxID=29461;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J., Van Aken S.B.,
 RA Riedmiller S., Tetrelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.E., Hailing S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 animal and plant pathogens and symbionts."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 DR EMBL; AE014350; AAN29340.1;
 KW TIGR; BR0394;
 KW Kinase; Complete proteome.
 SO SEQUENCE 299 AA; 31188 MW; A9B88F4157BF513 CRC64;

Query Match 20.7%; Score 313.5; DB 16; Length 299;
 Best Local Similarity 34.5%; Pred. No. 2.9e-21;
 Matches 92; Conservative 42; Mismatches 104; Indels 29; Gaps 11;
 QY 13 APATLMLRTGRRPPGTHLQTFVOMLDCMLTFHPVDDGRVTLRNPISGV-EGD 70
 DB 15 APAKINFLDTLFRKDDGHEIEVMVTVLNDRLSPERKDNKIVDIEHNVFNDKN 65
 QY 71 DLTVRAANILKSHHTGCVRGVCIIDIEKNLPMGGGLGGSSDAATLVVLRMLGSLSKRE 129
 DB 75 NLTKARDLRLRQHGPDLSPPVAIHLEKNLPISGIGGSSDAATLALNTLMQDLDFE 134
 QY 130 ELMDIGRLGADVVFVFGGSAMG-----EGVSEDLQATL-----LPEQWFTIIRKDCVNT 180
 DB 135 MAAATGLSLGADIPKCLHG-AAHGTPILARGLGEBELNDVSGIAALP---MLVNDGTALA 190
 QY 181 TGEIFSA-----ENLTNSAVVTMSDFLAG---DNRDCEVCKLYRPVKAIDALLCYA 233
 DB 191 TDPVFALTRRRRAFLPPRACGTDLCAYLRETRNDLPPAISIAPQIEPKLALRAKG 250
 QY 234 E--ARLTGTACVFAQPCNKEDAESAL 258
 DB 251 ALYAKMSGGATCFALFSD---ESAL 273
 RESULT 9
 Q8C0U6 PRELIMINARY; PRT; 282 AA.
 ID Q8C0U6
 AC Q8C0U6;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase.
 GN SE2288.
 OS Staphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 CX NCBI_TaxID=1282;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
 RA Chen Z., Wen Y.;
 RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AE016751; AAO05930.1;
 KW Kinase; Complete proteome.
 SO SEQUENCE 282 AA; 31030 MW; ASD58B0A01D48287 CRC64;
 Query Match 19.9%; Score 302; DB 16; Length 282;
 Best Local Similarity 32.0%; Pred. No. 3.2e-20;
 Matches 79; Conservative 51; Mismatches 103; Indels 14; Gaps 6;
 QY 13 APATLMLRTGRRPPGTHLQTFVOMLDCMLTFHPVDDGRVTLRNPISGV-EGD 71
 DB 6 APAKINFLDTLFRKDDGHEIEVMVTVLNDRLSPERKDNKIVDIEHNVFNDKN 65
 QY 72 LTVRAANILKSHHTGCVRGVCIIDIEKNLPMGGGLGGSSDAATLVVLRMLGSLSKRE 131
 DB 66 LAYKADLAFERENINBGTISIDKIPASALAGSADAATMGLENRLFELGSLDAL 125
 QY 133 MDIGRLGADVVFVFGGSAMG-----EGVSEDLQATL-----LPEQWFTIIRKDCVNTGEIFSAE 190
 DB 126 AALGIGIGIDIPFCYNGTAVCTGGEQVTFPKRPPSAMVVLAKENIGISSDVFRAIDL 185
 QY 191 TNSAV-----VTMSDF-----LAGDNRDCEVCKLYRPVKAIDALLCYA-ARLT 238
 DB 186 TEHHIVHEKCKQALENDYHLLCNLSLNRLPEPVSMAHPDIKKIKDNMLQCGADGALMS 245
 QY 239 GTGACVF 245
 DB 246 GSGPTVY 252

DB 212 LQMSGSGSDIPFFSQAAYCTGRGB-IVEDIHNPANLP---MVLKPEACSTAE 267

QY 184 IFS---AENLTRSAVYTMEDPLAGNRNDCSEVVC 216
DB 268 VYKRLRLHTSQTDPVLKELI---TENGISQDAC 299

RESULT 13

Q86Q24 PRELIMINARY; PRT; 315 AA.

AC Q86Q24; PRELIMINARY; PRT; 315 AA.
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Putative isopentenyl-monophosphate kinase.
GN C00973.

OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawabata Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RT Submitted (MAY/2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP005217; BAC17783.1; -
KM Kinase; Complete proteome.
SQ SEQUENCE 315 AA; 32871 MW; 472052CC7D843763 CRC64;

Query Match 14.7%; Score 223.5; DB 16; Length 315;
Best Local Similarity 28.3%; Pred. No. 9.6e-13;
Matches 80; Conservative 46; Mismatches 116; Indels 41; Gaps 12;

QY 4 AAGGGEWPAKALNEMLRITGRPPDGYHLLQTVFQMLDCMLTFHPVD----- 54
DB 5 ARAGG-----KINMLGVGPAKEDYHVLVTFQITLAEITITLTLEDELVEEGSV 57
QY 55 GRVTLRPISGV-EQDDLTVRANLPL-----KSHYGC-VKGVGIDIEKNLPMGGGLGGG 108
DB 58 RQILVTP-RGVPTTPNLMRAVDALVGRRRHRDRPLAVLHIDKIGFVAGGAGGS 116
QY 109 SDAATTLVLRNLGLSKRELMDLGLRGADVPVFGCSAMGEGVSEDLQATILP-2 166
DB 117 ADAAAALRAVD-AWIGPGEETILEVAALIGSDVPFCLGRTKLGTRGEQLVDMLSRGT 175
QY 167 EQMFVILKPPCHVVTGSIIFS---AENLTRSAVYTMEDPLAGNRNDCSEVVCILRPV- 222
DB 176 YHMLLVSPK-GLSTPEVFAKFDMSLPSSMDVTPMSQALLDSAGALAEVLNDLAPAA 234
QY 223 -----KDAIDALLCYAERLITGTGACVPAQFCNKEDAES 257
DB 235 LSLRPDLKXQLAGIRAGALATWVGSGPTIALLC--DDAQSA 275

RESULT 14

Q86G14 PRELIMINARY; PRT; 316 AA.

AC Q86G14; PRELIMINARY; PRT; 316 AA.
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase, Cmk,
DE 4 (cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase.
GN ISPG OR BL0656.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmaliantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwielen M.-C., Desiere F., Boik P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract";
RT Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).
DR EMBL: AE014687; AAN24478.1; -
KM Kinase; Complete proteome.
SQ SEQUENCE 316 AA; 33286 MW; AEF623697DB7875B CRC64;

Query Match 13.5%; Score 204.5; DB 16; Length 316;
Best Local Similarity 25.8%; Pred. No. 6e-11;
Matches 74; Conservative 44; Mismatches 118; Indels 51; Gaps 8;

QY 14 PAKMLMLRITGRPPD-GYHLLQTVFQMLDCMLTFHPVDGRTLRNPISGVPEQ-- 69
DB 30 PAKTMLTLEVGPAPHDEWGRHEDLTIYCAIGYDTVT-----ATAKQPAQFSLE 81
QY 70 -----DDLTVRANLTKSHRGCV-----RGVCIDIEKNLPMGGGLGGSSDAAT 114
DB 82 GAYLDGLASSRSDMRNRAVALPMAQAAREPVALITKRIYVAGAGGGSDAAT 141
QY 115 LVTLNRLGLSKRELMDLGLRGADVPVFGCSAMGEGVSEDL-----QATILP 166
DB 142 MLAVNRLMELNMPILRLTIAATLADMPFCLTGLAVGTGFRITDIAPGRDELALI 201
QY 167 EQMFVILKPPCHVVTGSIIFS---AENLTRSAVYTMEDPLAGNRNDCSEVVCILRPV- 222
DB 202 EQGF-----GSEVAVGAYGOSLTPPEYHTFTDIVGAEGGRNHLQAAALSLHPS 251
QY 223 KDAIDAL--LCYARLITGTGACVPAQFCNKEDAESLEGLKDRMLV 267
DB 255 GQAIIDAAQAGSHAFVSGSGSPSVAFADAEAAQRIIEVWRDTAVV 298

RESULT 15

Q8E2M8 PRELIMINARY; PRT; 297 AA.

AC Q8E2M8; PRELIMINARY; PRT; 297 AA.
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE 4-diphosphocytidylyl-2C-methyl-D-erythritol kinase (EC 2.7.1.-).
GN YCHB OR LA3824.

OS Leptospiro interogens.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE011537; AAN51022.1; -
KM Kinase; Transferase; Complete proteome.
SQ SEQUENCE 297 AA; 33271 MW; 51A1F018DAA7872 CRC64;

Query Match 12.9%; Score 196; DB 16; Length 297;
Best Local Similarity 26.3%; Pred. No. 3.5e-10;
Matches 77; Conservative 53; Mismatches 115; Indels 48; Gaps 11;

QY 13 AAKNLMRLIRGRPPDGYHLLQTVFQMLDCMLTFHPVDGRTLRNPISGVPEQ--RNP 62
DB 3 SPKNNLGLIETPKKLDGFHRSVFLKISMDDIRIPASGVVELSNNEILLKRL 62
QY 63 ISGVEQDGLT-----VRAANLTKSHYGCVRGVCIDIEKNLPMGGGLGGSSDAAT 114
DB 63 YQVSEKGIKNNILYKTFIKARSIFP---LPQKILHTRISIPAGLGGSTMAAS- 117
QY 115 LVTLNRL--WGLSKRELMDLGLRGADVPVFGCSAMGEGVSEDLQATILP-QMFV 171
DB 118 --LNFPLSPMRPFSTSDMFVLAETIGSDVPFCLGRTKLGTRGEQLVDMLSRGT 175

QY	172	IIPDCHVNTGELIS-----ABNLTRN-SAVVTMSDF--LAGDRNRDCSEV	214
Db	176	ALTPQV-NTISEMYSLIKPLQESASQKNGNTLSKNLISILKNGDWSLQGLANDPEPV	234
QY	215	VCKLYRPVKDAIDALLCYAEA--RLNGTACVFAQFCNKEDAESALEGLKDRW	265
Db	235	AFQHPPELAGVLKDKLFEGSSYCSLNGSSSMYGLVQGLEIQEELLQRLRQEF	287

Search completed: January 29, 2004, 15:54:08
 Job time : 23.2436 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:40:33 Search time 14.3511 Seconds
(without alignments)
1736.455 Million cell updates/sec

Title: US-09-941-947a-14

Perfect score: 806

Sequence: 1 MIRXGCVDFHRFNDGDHIL.....LGFGRKRGIAVAVVLIEN 157

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	806	100.0	157	23	ABG61583
2	806	100.0	157	23	AAH22305
3	806	100.0	157	23	AAH22305
4	537	66.6	159	21	AAH11373
5	537	66.6	159	21	AAH27149
6	537	66.6	159	22	AAH27149
7	537	66.6	159	24	ABP70685
8	534	66.3	158	24	ABP70685
9	534	66.3	158	21	AAH27151

10	531	65.9	159	24	ABP70686
11	528	65.5	159	24	ABP70687
12	528	65.5	159	24	ABP70688
13	519	64.4	170	24	ABG75686
14	507	62.9	174	24	ABP79612
15	505	62.7	160	24	ABU06061
16	451	56.0	157	23	ABH49155
17	438	54.3	191	22	ABG18263
18	405	50.2	120	23	ABP02539
19	396	49.1	161	21	AAH27152
20	377	46.8	240	21	AAH15275
21	377	46.8	240	21	AAH15275
22	374	46.4	231	21	AAH27449
23	374	46.4	231	21	AAH27448
24	353	43.8	259	21	AAH27633
25	345	42.8	409	19	AAH11013
26	343	42.6	406	21	AAH27154
27	341	42.3	352	18	AAH20628
28	329	40.8	399	21	AAH27153
29	321	39.8	223	23	ABH93990
30	286	35.5	176	22	AAH50581
31	269	33.4	240	21	AAH18226
32	269	33.4	248	22	AAH70488
33	247.5	30.7	158	24	AAH31686
34	242	30.0	161	21	AAH39032
35	229	28.4	174	23	ABH65843
36	213.5	26.5	160	22	AAH92667
37	213.5	26.5	160	22	AAH80182
38	153	19.0	199	20	AAH35167
39	151	18.7	96	18	AAH20173
40	117.5	14.6	102	20	AAH37797
41	86	10.7	434	21	AAH08216
42	83.5	10.4	410	23	ABP28012
43	82	10.2	943	22	AAH91445
44	81	10.0	434	21	AAH38717
45	81	10.0	434	23	ABH92639

ALIGNMENTS

RESULT 1	ABG61583	High growth mechanotrophic bacterial strain polypeptide #33.
ID	ABG61583	standard; Protein; 157 AA.
AC	ABG61583	
XX	XX	
DT	27-AUG-2002 (first entry)	
XX	XX	
DE	High growth mechanotrophic bacterial strain polypeptide #33.	
XX	XX	
KM	High growth mechanotrophic bacterial strain; C1 carbon substrate; enzyme; methane; methanol; Embden-Meyerhof carbon flux pathway; 16S rRNA;	
KM	hydrophosphate dependent phosphofructokinase; nitrogen-containing compound;	
KM	ammonia; nitrate; nitrite; nitrogen; pigment; oxygen; landfill;	
KM	methane-containing environment; waste water treatment system; isoprenoid;	
KM	nitrous oxide; terpenoid; animal feed; carotenoid; exopolysaccharide.	
XX	XX	
OS	Methylobionas 16a.	
XX	XX	
FN	WO200220728-A2.	
XX	XX	
PD	14-MAR-2002.	
XX	XX	
PF	28-AUG-2001; 2001WO-US26827.	
XX	XX	
PR	01-SEP-2000; 2000US-229858P.	
XX	XX	
PA	(DUPO) DU PONT DE NEMOURS & CO E. I.	
XX	XX	
PI	Koffas M, Odem JM, Schenzle A;	
XX	XX	
DR	WPI; 2002-452200/48.	

2C-methyl-D-erythr
2C-methyl-D-erythr
2C-methyl-D-erythr
2C-methyl-D-erythr
N. gonorrhoeae and
N. meningitidis va
Listeria monocyco
Novel human diagno
Human ORF protein
Synchocystis esse
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
H. pylori ORF hpt
H. pylori ORF hpt
H. pylori cytoplasm
T. pallidum essent
Herbicidally activ
Propionibacterium
Plasmodium falcipar
Plasmodium falcipar
Rhodococcus erythr
Arabidopsis thalia
Bifidobacterium 10
C glutamicum prote
Corynebacterium 91
Chlamydia pneumonia
H. pylori cytoplasm
Amino acid sequenc
Arabidopsis thalia
Streptococcus poly
C glutamicum prote
Arabidopsis thalia
Herbicidally activ

DR N-PSDB; ABK63262.
 XX
 PT New high growth methanotrophic bacterial strain, useful for producing
 PT single cell proteins, grows on a C1 carbon substrate, and comprises a
 PT functional gene encoding in Embden-Meyerhof carbon pathway
 PS Claim 11; Page 142-143; 157pp; English.
 CC The invention relates to a high growth methanotrophic bacterial strain,
 CC which grows on a C1 carbon substrate e.g. methane and methanol, and
 CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a
 CC gene coding a pyrophosphate dependent phosphofructokinase enzyme or a 16S
 CC rRNA. The bacterial strain is useful for the production of single cell
 CC protein and for the biotransformation of a nitrogen-containing compound,
 CC e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the
 CC production of a feed product comprising a protein, carbohydrates and a
 CC pigment and for reducing oxygen demand, for removing nitrates and
 CC nitrates in methane-containing environments such as landfills, waste
 CC water treatment systems or anywhere that methane, oxygen and nitrates are
 CC present. The bacterial strain of the invention can be used as a
 CC denitrifying agent for the conversion of nitrate or nitrite to nitrous
 CC oxide with methane or methanol as a carbon source. It is also used in the
 CC production of biomass including proteins, carbohydrates and a wide
 CC variety of pigments (particularly for isoprenoid pigments for the
 CC purpose of generating animal feeds), in production of terpenoid and
 CC carotenoid compounds, useful as pigments and as monomers in polymeric
 CC materials and in production of exopolysaccharides at high levels.
 CC Sequences ABG61551-ABG61590 represent high growth methanotrophic
 CC bacterial strain proteins of the invention.
 XX
 SQ Sequence 157 AA;
 Query Match 100.0%; Score 806; DB 23; Length 157;
 Best Local Similarity 100.0%; Pred. No. 1.2e-87;
 Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MIRVGAGYVHRNDGDHIIIGGVKIPYKGLFAHSDGVVLAALADALIGAAALGDIGK 60
 DB 1 MIRVGAGYVHRNDGDHIIIGGVKIPYKGLFAHSDGVVLAALADALIGAAALGDIGK 60
 QY 61 HPPDTPNFKGADSRVLLRHVYGVIVKKGKYLVAADVITIIAQAQKMLPHVPGKRAANIAAD 120
 DB 61 HPPDTPNFKGADSRVLLRHVYGVIVKKGKYLVAADVITIIAQAQKMLPHVPGKRAANIAAD 120
 QY 121 LETDVDFINVKATTTEKLGFEGRKEGIAVQAVVLIIR 157
 DB 121 LETDVDFINVKATTTEKLGFEGRKEGIAVQAVVLIIR 157
 RESULT 2
 AAE22305
 ID AAE22305 standard; Protein; 157 AA.
 XX
 AC AAE22305;
 XX
 DT 25-JUL-2002 (first entry)
 XX
 DE Methylomonas 16a 2C-methyl-d-erythritol 2,4-cyclodiphosphate synthase.
 XX
 KW Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astatxanthin; diet;
 KW anti-oxidant; steroid; fragrance; electro-optic application;
 KW aquaculture; 2C-methyl-d-erythritol 2,4-cyclodiphosphate synthase; ispF;
 KW enzyme.
 XX
 OS Methylomonas 16a sp.
 XX
 PN WO200218617-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 04-SEP-2001; 2001WO-US27420.
 XX
 PR 01-SEP-2000; 2000US-229858P.

PR 01-SEP-2000; 2000US-229907P.
 XX
 PA (DUPO) DU POINT DE MEMOIRS & CO B I.
 XX
 PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;
 PI Odum JM, Picataggio SK, Rouviere PB;
 XX
 DR WPI: 2002-351711/38.
 DR N-PSDB; AAD35503.
 XX
 PT Producing carotenoid compounds e.g. antheraxanthin and astatxanthin, by
 PT using microorganisms having a nucleic acid molecule encoding enzymes in
 PT the carotenoid biosynthetic pathway and which metabolize single carbon
 PT substrates
 PS Claim 45; Page 118-119; 156pp; English.
 CC The invention relates to a method for producing carotenoid compounds.
 CC The method comprises a transformed metabolizing host cell, comprising
 CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule
 CC encoding an enzyme in the carotenoid biosynthetic pathway, under the
 CC control of regulatory sequences, and contacting the host cell with carbon
 CC substrate to produce a carotenoid compound. The method is useful for
 CC producing carotenoid compounds such as antheraxanthin and astatxanthin, by
 CC using microorganism having a nucleic acid molecule encoding enzymes in
 CC the carotenoid biosynthetic pathway and which metabolize single carbon
 CC substrates. The carotenoid have potent anti-oxidant properties useful in
 CC diet, and aquaculture elements. The carotenoids are also useful as
 CC intermediates in the synthesis of steroids flavours and fragrances and
 CC compounds for potential electro-optic applications. The present sequence
 CC is Methylomonas 16a sp. 2C-methyl-d-erythritol 2,4-cyclodiphosphate
 CC synthase (ispF) enzyme used in the invention.
 XX
 SQ Sequence 157 AA;
 Query Match 100.0%; Score 806; DB 23; Length 157;
 Best Local Similarity 100.0%; Pred. No. 1.2e-87;
 Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MIRVGAGYVHRNDGDHIIIGGVKIPYKGLFAHSDGVVLAALADALIGAAALGDIGK 60
 DB 1 MIRVGAGYVHRNDGDHIIIGGVKIPYKGLFAHSDGVVLAALADALIGAAALGDIGK 60
 QY 61 HPPDTPNFKGADSRVLLRHVYGVIVKKGKYLVAADVITIIAQAQKMLPHVPGKRAANIAAD 120
 DB 61 HPPDTPNFKGADSRVLLRHVYGVIVKKGKYLVAADVITIIAQAQKMLPHVPGKRAANIAAD 120
 QY 121 LETDVDFINVKATTTEKLGFEGRKEGIAVQAVVLIIR 157
 DB 121 LETDVDFINVKATTTEKLGFEGRKEGIAVQAVVLIIR 157
 RESULT 3
 AAU80329
 ID AAU80329 standard; Protein; 157 AA.
 XX
 AC AAU80329;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Methylomonas 16a ORF5 ygbB/ispF protein sequence.
 XX
 KW Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;
 KW keratinoid; pigment; flavour; fragrance; open reading frame 5; ORF5;
 KW ygbB; 2C-methyl-d-erythritol 2,4-cyclodiphosphate synthase enzyme; ispF.
 XX
 OS Methylomonas sp.
 XX
 PN WO200220733-A2.
 XX
 PD 14-MAR-2002.
 XX
 PR 29-AUG-2001; 2001WO-US26852.

XX 01-SEP-2000; 2000US-229907P.
 PR (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX
 PA
 PI Cheng Q, Kofias M, Norton KC, Odom JM, Picataggio SK, Rouviere PE,
 PI Schenzle A, Tomb J;
 XX
 XX WPI: 2002-383051/41.
 DR N-PSDB; AAK50085.
 XX
 PT Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,
 PT isolated from *Methylobionas 16a*, useful for the production of isoprenoid
 PT compounds -
 XX
 PS Claim 4; Page 73; 84pp; English.
 XX
 CC The present invention relates to a new nucleic acid molecule encoding
 CC an isoprenoid biosynthetic enzyme isolated from *Methylobionas 16a*.
 CC The invention is useful for obtaining a nucleic acid molecule
 CC encoding an isoprenoid compound biosynthetic enzyme, and for the
 CC microbial production of isoprenoid compounds. The molecules of the
 CC invention are also useful for regulating isoprenoid biosynthesis in an
 CC organism and for producing recombinant organisms for producing various
 CC isoprenoid compounds. The nucleic acid is also useful for feed additive,
 CC for the production of ketonoids and their derivatives, isoprenoid
 CC intermediates, and as pure products useful as pigments, flavours and
 CC fragrances. The present amino acid sequence represents the *Methylobionas*
 CC 16a open reading frame 5 (ORF5) ygbB/ispF (2C-methyl-d-erythritol
 CC 2,4-cyclodiphosphate synthase enzyme) protein of the invention, as
 CC described above.
 CC
 SQ Sequence 157 AA;
 XX
 XX
 Query Match 100.0%; Score 806; DB 23; Length 157;
 Best Local Similarity 100.0%; Pred. No. 1.2e-87;
 Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 MIRVGAGYDVHRRNDGHIILGCVKIPYKGLBAHSDGVVLAHALDAILGAALADIGK 60
 DB 1 MIRVGAGYDVHRRNDGHIILGCVKIPYKGLBAHSDGVVLAHALDAILGAALADIGK 60
 QY 61 HEPDTPNPFAGDSRYLIRHYGIVVEKGYKLVNADVTIIAQPXKLPHVPGKRANIAD 120
 DB 61 HEPDTPNPFAGDSRYLIRHYGIVVEKGYKLVNADVTIIAQPXKLPHVPGKRANIAD 120
 QY 121 LETDVPFINVKATTTTEKLGFEGRKEGIAVOAVVLI 157
 DB 121 LETDVPFINVKATTTTEKLGFEGRKEGIAVOAVVLI 157
 XX
 RESULT 4
 AAB11373
 ID AAB11373 standard; protein; 159 AA.
 XX
 AC AAB11373;
 XX
 DT 22-FEB-2001 (first entry)
 XX
 DE E. coli YGBB protein.
 XX
 DE E. coli YGBB protein.
 XX
 KW YGBB; KOTB; YOGF; YHBC; YGBB; YGBB; YCHB; antibacterial;
 KW treatment; infection.
 XX
 OS *Escherichia coli*.
 XX
 PN DE19916176-A1.
 XX
 PD 12-OCT-2000.
 XX
 PF 10-APR-1999; 99DE-1016176.
 XX
 PR 10-APR-1999; 99DE-1016176.

XX (FARB) BAYER AG.
 PA
 PI Broetz H, Ehler K, Freiberg C, Spaltmann F, Wieland B;
 PI Labischinski H;
 XX
 XX WPI: 2000-639611/62.
 DR N-PSDB; AAC66043.
 XX
 PT Essential genes from bacteria, useful in screening for antimicrobial
 PT agents, and related proteins, transformants and antisense sequences -
 XX
 PS Disclosure; Page 22-23; 28pp; German.
 XX
 CC This invention describes novel *Escherichia coli* genes (I) encoding
 CC proteins (II) designated YOGF, YHBC, YGCU, YGBB, YCHB, YGBB, YGBB and
 CC KDTB, and genes (Ia) that encode orthologous gene products (Iia) in
 CC other microorganisms and which have antibacterial activity. Recombinant
 CC microorganisms in which expression of (I) or (Ia) can be regulated are
 CC used to identify compounds that bind to the gene products, particularly
 CC in affinity selection assays. (II) and (Iia) are used to identify, or
 CC prepare, antibodies and other proteins that bind to the gene products.
 CC Substances that bind to (II) or (Iia) are potentially useful as
 CC antibacterials for treating a wide range of infections in humans and
 CC animals. Sequences antisense to (I) and (Ia) can also be used as
 CC antibacterials. The specified genes are widely distributed in bacteria
 CC but have no close homologs in eukaryotic cells.
 CC
 SQ Sequence 159 AA;
 XX
 XX
 Query Match 66.6%; Score 537; DB 21; Length 159;
 Best Local Similarity 69.5%; Pred. No. 1.4e-55;
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;
 XX
 QY 2 IRVAGYDVHRRNDGHIILGCVKIPYKGLBAHSDGVVLAHALDAILGAALADIGK 61
 DB 1 MIRVGAGYDVHRRNDGHIILGCVKIPYKGLBAHSDGVVLAHALDAILGAALADIGK 60
 QY 62 FPDTPNPFAGDSRYLIRHYGIVVEKGYKLVNADVTIIAQPXKLPHVPGKRANIAD 121
 DB 61 FPDTPNPFAGDSRYLIRHYGIVVEKGYKLVNADVTIIAQPXKLPHVPGKRANIAD 120
 QY 122 ETDVPFINVKATTTTEKLGFEGRKEGIAVOAVVLI 155
 DB 121 GCHMDVNVKATTTTEKLGFEGRKEGIAVOAVVLI 154
 XX
 RESULT 5
 AAB27149
 ID AAB27149 standard; Protein; 159 AA.
 XX
 AC AAB27149;
 XX
 DT 27-FEB-2001 (first entry)
 XX
 DE E. coli essential protein ygbB.
 XX
 DE Bacterial growth; inhibitor; ygbB; yfHC; yACE; yCHB; yEJD; yfH;
 KW YGCU; YGBB; YGCU; yfHC; yHBC; yGBB; yGBB; yGBB; yGBB; yGBB;
 KW b1808; yGBB; yGBB; yGBB; yGBB; yGBB; yGBB; yGBB; yGBB; yGBB;
 KW H. influenza; S. pneumoniae; H. pylori; B. burgdorferi; pesticide;
 KW M. tuberculosis; antibiotic.
 XX
 OS *Escherichia coli*.
 XX
 PN WO200061793-A2.
 XX
 PD 19-OCT-2000.
 XX
 PF 07-APR-2000; 2000MO-BP03135.
 XX
 PR 09-APR-1999; 99EP-0107031.
 XX
 PR 04-FEB-2000; 2000EP-0102111.

XX (GPCB-) GPC BIOTECH AG.
 PA Loferer H, Jacobi A;
 PI WPI; 2000-687048/67.
 DR
 XX Identifying antibacterial compounds, comprises identifying an
 PT antagonist or inhibitor of the expression of a gene encoding a
 PT polypeptide essential for bacterial growth or survival -
 PS Disclosure; Fig 4; 75bp; English.
 XX
 CC The present invention relates to antagonists and inhibitors of 24
 CC bacterial genes and proteins. The proteins are thought to be essential
 CC for growth in several species of bacteria (including *S. pneumoniae*, *B.*
 CC *burgorferi*, *H. influenza* and *H. pylori*). The proteins and coding
 CC sequences shown in the specification can be used to identify antagonists
 CC and inhibitors which can be used in disease treatment and pesticides. In
 CC particular, they can be used against *M. tuberculosis*. The present
 CC sequence is one of the proteins of the invention.
 CC
 SQ Sequence 159 AA;
 Query Match 66.6%; Score 537; DB 21; Length 159;
 Best Local Similarity 69.5%; Pred. No. 1.4e-55;
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;
 QY 2 IRVNGYDVERFENDGHIILGGVYKIPYKGLBAHSDGVDVLAHALADALIGAAALGDIQKH 61
 DB 1 MRIGHGFVHAFGGEGPILIGVYRIPEKGLAHSDGVDVLAHALADALIGAAALGDIQKL 60
 QY 62 PPDDPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQA PKLPVHVGKRAMIAADL 121
 DB 61 PPDDPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQA PKLPVHVGKRAMIAADL 120
 QY 122 EFDVDFINRKATTEKLGFEGRKEGIAVQAVVLI 155
 DB 121 GCHMDVNVKATTEKLGFEGRKEGIAVQAVVLI 154
 RESULT 6
 AAG63811
 ID AAG63811 standard; Protein; 159 AA.
 AC AAG63811;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthetase.
 XX
 KM 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthetase;
 KM non-mevalonate pathway; 2-C-methyl-D-erythritol-2,4-cyclodiphosphate;
 KM 2-phospho-4-(cytidine 5'-diphospho) 2-C-methyl-D-erythritol; CDP-ME2P;
 KM cytidine 5'-monophosphate; isoprenoid; antibacterial.
 OS
 XX *Escherichia coli*.
 XX
 PN WO200157223-A1.
 XX
 PD 09-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-JP00483.
 XX
 PR 03-FEB-2000; 2000US-0180126.
 PR 07-FEB-2000; 2000JP-0029287.
 XX
 PA (SETO/) SETO H.
 PA (KUZUYAMA T. KUZUYAMA T.
 XX
 PI Seto H, Kuzuyama T;
 XX
 DR WPI; 2001-522353/57.

DR N-PSDB; AAR74788.
 XX
 PT 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthetase, useful for
 PT producing isoprenoids and for screening for non-mevalonate pathway
 PT inhibitors -
 XX
 PS Claim 2; Page 42-43; 49pp; Japanese.
 XX
 CC The present sequence represents a
 CC 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthetase. The enzyme is
 CC involved in the non-mevalonate pathway. It acts on
 CC 2-phospho-4-(cytidine 5'-diphospho) 2-C-methyl-D-erythritol (CDP-ME2P) -
 CC to produce 2-C-methyl-D-erythritol-2,4-cyclodiphosphate and cytidine
 CC 5'-monophosphate. The enzyme has a molecular weight of 22 kDa, as
 CC determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis.
 CC The enzyme or gene can be used to increase isoprenoid compound
 CC production, and also for screening for antibacterials.
 CC
 SQ Sequence 159 AA;
 Query Match 66.6%; Score 537; DB 22; Length 159;
 Best Local Similarity 69.5%; Pred. No. 1.4e-55;
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;
 QY 2 IRVNGYDVERFENDGHIILGGVYKIPYKGLBAHSDGVDVLAHALADALIGAAALGDIQKH 61
 DB 1 MRIGHGFVHAFGGEGPILIGVYRIPEKGLAHSDGVDVLAHALADALIGAAALGDIQKL 60
 QY 62 PPDDPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQA PKLPVHVGKRAMIAADL 121
 DB 61 PPDDPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQA PKLPVHVGKRAMIAADL 120
 QY 122 EFDVDFINRKATTEKLGFEGRKEGIAVQAVVLI 155
 DB 121 GCHMDVNVKATTEKLGFEGRKEGIAVQAVVLI 154
 RESULT 7
 ABP70685
 ID ABP70685 standard; Protein; 159 AA.
 AC ABP70685;
 XX
 DT 04-APR-2003 (first entry)
 XX
 DE *Escherichia coli* 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase.
 XX
 KM Protein co-ordinate data; herbicide; protozoacide; antibacterial;
 KM antibiotic; isoprenoid synthesis; crystal; enzyme;
 KM 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; malaria;
 KM Plasmodium falciparum; ispf.
 OS
 XX *Escherichia coli*.
 XX
 PN WO200292800-A2.
 XX
 PD 21-NOV-2002.
 XX
 PF 13-MAY-2002; 2002WO-BP05238.
 XX
 PR 15-MAY-2001; 2001DE-1023597.
 PR 25-MAY-2001; 2001US-293875P.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (BACHER) BACHER A.
 XX
 PI Bacher A, Hecht S, Huber R, Kaiser J, Rohdich F, Steinbacher S;
 XX
 DR WPI; 2003-167170/16.
 DR N-PSDB; AAZ25497.
 XX
 PT Novel crystal useful for identifying a potential inhibitor of
 PT 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, comprises

PT 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase protein with or
 PT without zinc
 XX
 PS Example 2; Fig 9; 21app; English.
 CC The present invention relates to a crystal comprising
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase with or without
 CC zinc. The crystal is useful in an inhibitor-screening assay, and
 CC identified inhibitors are useful as antibiotics against bacteria or
 CC protozoa especially malarial parasite e.g. Plasmodium falciparum, and
 CC as herbicides. The inhibitors are also useful for inhibiting the
 CC synthesis of isoprenoids. The present sequence is 2C-methyl-D-erythritol
 CC 2,4-cyclodiphosphate synthase from Escherichia coli, encoded by the *ispS*
 CC gene.
 XX
 SQ Sequence 159 AA;
 Query Match 66.6%; Score 537; DB 24; Length 159;
 Best Local Similarity 69.5%; Pred. No. 1.4e-55;
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;
 QY 2 IRVGMGVDRFNRDGHILIGVKIPYKGLAEHSDGVYTHALADAILGAAAGDIGK 61
 DB 1 MRIGHGSDVHAFGEGPILIGVRIPEKGLAHSDGVYTHALADAILGAAAGDIGK 60
 QY 62 FPDTPNFKGADSRVLLRHVYGIKGYLGNADVTIIAQAPKMLPHVPGKRNIAADL 121
 DB 61 FPDTPNFKGADSRVLLRHVYGIKGYLGNADVTIIAQAPKMLPHVPGKRNIAADL 120
 QY 122 ETVDDPFINVATTTETKLGFBGKKGIAVQAVLTIER 155
 DB 121 GCHMDVNVVATTTETKLGFBGKKGIAVQAVLTIER 154
 DB
 RESULT 8
 AAB27150
 ID AAB27150 standard; Protein; 158 AA.
 AC AAB27150;
 XX
 XX 27-FEB-2001 (first entry)
 DE H. influenzae essential protein ygbB.
 XX
 XX Bacterial growth; inhibitor; ygbB; yfnc; yacE; ychB; yebD; yfii;
 KM yggJ; yjeb; yjao; yrdC; yhbc; ygbP; ybeY; gcbE; kdtB; pfs; ycaJ;
 KM b1808; yeaA; yagF; b1983; yldD; yceg; yjbc; antibacterial compound;
 KM H. influenza; S. pneumoniae; H. pylori; B. burgdorferi; pesticide;
 KM M. tuberculosis; antibiotic.
 XX
 XX Haemophilus influenzae.
 OS
 XX WO200061793-A2.
 PN
 XX 19-OCT-2000.
 PD
 XX 07-APR-2000; 2000WO-3P03135.
 PF
 XX 09-APR-1999; 99EP-0107031.
 PR
 XX 04-FEB-2000; 2000EP-0102111.
 XX
 PA (GPCB-) GPC BIOTECH AG.
 PI
 XX Loferer H, Jacobi A;
 PT
 XX WPI; 2000-687048/67.
 DR
 XX
 XX Identifying antibacterial compounds, comprises identifying an
 PT antagonist or inhibitor of the expression of a gene encoding a
 PT polypeptide essential for bacterial growth or survival -
 PS Disclosure; Fig 4; 75pp; English.
 XX

CC The present invention relates to antagonists and inhibitors of 24
 CC bacterial genes and proteins. The proteins are thought to be essential
 CC for growth in several species of bacteria (including S. pneumoniae, B.
 CC burgdorferi, H. influenza and H. pylori). The proteins and coding
 CC sequences shown in the specification can be used to identify antagonists
 CC and inhibitors which can be used in disease treatment and pesticides. In
 CC particular, they can be used against M. tuberculosis. The present
 CC sequence is one of the proteins of the invention.
 XX
 SQ Sequence 158 AA;
 Query Match 66.3%; Score 534; DB 21; Length 158;
 Best Local Similarity 65.6%; Pred. No. 3.1e-55;
 Matches 103; Conservative 20; Mismatches 34; Indels 0; Gaps 0;
 QY 1 MTRVGMGVDRFNRDGHILIGVKIPYKGLAEHSDGVYTHALADAILGAAAGDIGK 60
 DB 1 MRIGHGSDVHAFGEGPILIGVRIPEKGLAHSDGVYTHALADAILGAAAGDIGK 60
 QY 61 HPPDTPNFKGADSRVLLRHVYGIKGYLGNADVTIIAQAPKMLPHVPGKRNIAAD 120
 DB 61 LFPDTPNFKGADSRVLLRHVYGIKGYLGNADVTIIAQAPKMLPHVPGKRNIAAD 120
 QY 121 LETDVPFINVATTTETKLGFBGKKGIAVQAVLTIER 157
 DB 121 LQCDIEQVNVVATTTETKLGFBGKKGIAVQAVLTIER 157
 DB
 RESULT 9
 AAB27151
 ID AAB27151 standard; Protein; 158 AA.
 AC AAB27151;
 XX
 XX 27-FEB-2001 (first entry)
 DE B. subtilis essential protein ygbB.
 XX
 XX Bacterial growth; inhibitor; ygbB; yfnc; yacE; ychB; yebD; yfii;
 KM yggJ; yjeb; yjao; yrdC; yhbc; ygbP; ybeY; gcbE; kdtB; pfs; ycaJ;
 KM b1808; yeaA; yagF; b1983; yldD; yceg; yjbc; antibacterial compound;
 KM H. influenza; S. pneumoniae; H. pylori; B. burgdorferi; pesticide;
 KM M. tuberculosis; antibiotic.
 XX
 XX Bacillus subtilis.
 OS
 XX WO200061793-A2.
 PN
 XX 19-OCT-2000.
 PD
 XX 07-APR-2000; 2000WO-EP03135.
 PF
 XX 09-APR-1999; 99EP-0107031.
 PR
 XX 04-FEB-2000; 2000EP-0102111.
 XX
 PA (GPCB-) GPC BIOTECH AG.
 PI
 XX Loferer H, Jacobi A;
 PT
 XX WPI; 2000-687048/67.
 DR
 XX
 XX Identifying antibacterial compounds, comprises identifying an
 PT antagonist or inhibitor of the expression of a gene encoding a
 PT polypeptide essential for bacterial growth or survival -
 PS Disclosure; Fig 4; 75pp; English.
 XX
 XX The present invention relates to antagonists and inhibitors of 24
 CC bacterial genes and proteins. The proteins are thought to be essential
 CC for growth in several species of bacteria (including S. pneumoniae, B.
 CC burgdorferi, H. influenza and H. pylori). The proteins and coding
 CC sequences shown in the specification can be used to identify antagonists
 CC and inhibitors which can be used in disease treatment and pesticides. In
 CC particular, they can be used against M. tuberculosis. The present
 CC sequence is one of the proteins of the invention.

CC particular, they can be used against *M. tuberculosis*. The present
 CC sequence is one of the proteins of the invention.

XX Sequence 158 AA;

Query Match 66.3%; Score 534; DB 21; Length 158;
 Best Local Similarity 65.6%; Pred. No. 3.1e-55;

Matches 103; Conservative 18; Mismatches 36; Indels 0; Gaps 0;

QY 1 MIVGNGYVHRNDGHHILGKVPYKGLFAHSDGVVLAALADALGAALADIGK 60
 DB 1 MFRIGGFVHAFGSGPILIGGVRIPYKGLFAHSDGVVLAALADALGAALADIGK 60

QY 61 HEPDTPNFKGADSRVLAHVGIVKKGKYL VNA DVTIIAQP KKLPHVPGKRIANIAD 120
 DB 61 HEPDTPNFKGADSRVLAHVGIVKKGKYL VNA DVTIIAQP KKLPHVPGKRIANIAD 120

QY 121 LETVDVFINVKATTTKLGFEGRKKGIAVQAVVLI 157
 DB 121 LEADVGVNKKATTTKLGFEGRKKGIAVQAVVLI 157

RESULT 10
 ID ABP70686 standard; Protein; 159 AA.
 AC ABP70686;
 XX
 DT 04-APR-2003 (first entry)

DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase mutant DBS.

XX Protein co-ordinate data; herbicide; protozoacide; antibacterial;
 KM antibiotic; isoprenoid synthesis; crystal; enzyme;

KM 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; malaria;
 KM Plasmodium falciparum; ispf; mutant; mutein.

XX Escherichia coli.
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 8 /note= "Wild-type Asp substituted by Ser"

PN WO200292800-A2.

PD 21-NOV-2002.

PE 13-MAY-2002; 2002WO-EP05238.

PR 15-MAY-2001; 2001DE-1023597.

PR 25-MAY-2001; 2001US-293875P.

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PA (BACH/) BACHER A.

PI Bacher A, Hecht S, Huber R, Kaiser J, Rohdich F, Steinbacher S;

DR WPI; 2003-167170/16.

PT Novel crystal useful for identifying a potential inhibitor of
 PT 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, comprises
 PT 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase protein with or
 PT without zinc

XX Example 2; Page -; 214pp; English.

XX The present invention relates to a crystal comprising
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase with or without
 CC zinc. The crystal is useful in an inhibitor-screening assay, and
 CC identified inhibitors are useful as antibiotics against bacteria or
 CC protozoa especially malarial parasite e.g. Plasmodium falciparum, and
 CC as herbicides. The inhibitors are also useful for inhibiting the

CC synthesis of isoprenoids. The present sequence is a mutant of
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase which was generated
 CC in an example from the invention.

CC Note: The present sequence is not shown in the specification, but is
 CC derived from the wild-type sequence of Escherichia coli

CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ABP70685).

XX Sequence 159 AA;

Query Match 65.9%; Score 511; DB 24; Length 159;
 Best Local Similarity 68.8%; Pred. No. 7.2e-55;
 Matches 106; Conservative 14; Mismatches 34; Indels 0; Gaps 0;

QY 2 IRVNGYVHRNDGHHILGKVPYKGLFAHSDGVVLAALADALGAALADIGK 61
 DB 1 MFRIGGFVHAFGSGPILIGGVRIPYKGLFAHSDGVVLAALADALGAALADIGK 60

QY 62 PPDTPNFKGADSRVLAHVGIVKKGKYL VNA DVTIIAQP KKLPHVPGKRIANIAD 121
 DB 61 PPDTPNFKGADSRVLAHVGIVKKGKYL VNA DVTIIAQP KKLPHVPGKRIANIAD 120

QY 122 ETDVDVFINVKATTTKLGFEGRKKGIAVQAVVLI 155
 DB 121 GCHMDVNVKATTTKLGFEGRKKGIAVQAVVLI 154

RESULT 11
 ID ABP70687 standard; Protein; 159 AA.
 AC ABP70687;
 XX
 DT 04-APR-2003 (first entry)

DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase mutant H108.

XX Protein co-ordinate data; herbicide; protozoacide; antibacterial;
 KM antibiotic; isoprenoid synthesis; crystal; enzyme;

KM 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; malaria;
 KM Plasmodium falciparum; ispf; mutant; mutein.

XX Escherichia coli.
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 10 /note= "Wild-type His substituted by Ser"

PN WO200292800-A2.

PD 21-NOV-2002.

PE 13-MAY-2002; 2002WO-EP05238.

PR 15-MAY-2001; 2001DE-1023597.

PR 25-MAY-2001; 2001US-293875P.

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PA (BACH/) BACHER A.

PI Bacher A, Hecht S, Huber R, Kaiser J, Rohdich F, Steinbacher S;

DR WPI; 2003-167170/16.

PT Novel crystal useful for identifying a potential inhibitor of
 PT 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, comprises
 PT 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase protein with or
 PT without zinc

XX Example 3; Page -; 214pp; English.

XX The present invention relates to a crystal comprising
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase with or without

PT Novel 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase protein or
 PT its functional protein subunit, in crystalline form, useful for
 PT identifying and designing inhibitors and activators of the protein
 XX
 XX
 PS Disclosure; Fig 2; 372pp; English.

CC The invention relates to a 2C-methyl-D-erythritol 2,4-cyclodiphosphate
 CC synthase (MECS) protein, or a functional MECS protein subunit, in
 CC crystalline form. The invention also relates to a computer-readable
 CC database comprising representation of a compound capable of binding a
 CC binding pocket of an MECS protein. The methods are useful for producing
 CC a compound comprising a 3D molecular structure represented by the
 CC coordinates contained in a computer readable database, modulating MECS
 CC protein activity by contacting the MECS with a compound, identifying an
 CC activator or inhibitor of a protein that comprises an MECS active site
 CC or binding pocket, producing a mutant MECS protein having an altered
 CC property relative to an MECS protein and determining whether a compound
 CC binds the MECS protein. The protein is useful for identifying and
 CC designing inhibitors and activators of MECS, for designing
 CC anti-microbials that target the active site or a binding format of MECS
 CC or otherwise interfere with MECS activity or another activity in an
 CC associated biochemical, metabolic or anabolic pathway, or for rational
 CC drug design to identify and/or design compounds that bind MECS for
 CC developing new therapeutic agents. This sequence represents the MECS
 CC protein of the invention.

XX
 XX
 SQ Sequence 170 AA;

Query Match 64.4%; Score 519; DB 24; Length 170;
 Best Local Similarity 63.7%; Pred. No. 2.1e-53;
 Matches 100; Conservative 21; Mismatches 36; Indels 0; Gaps 0;

QY 1 MTRVGMGVDRHFRNCDHIIIGVYKIPYKGLFAHSDGDVYTHALADALIGAALGDIGK 60
 DB 3 LIRIGGFVHAFGEGRPLIGVEVPHYTGFIASHDGVNLAHALTDALIGAALGDIGK 62
 QY 61 HRPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQAAPKLPVPGKANIADL 120
 DB 63 LRPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQAAPKLPVPGKANIADL 122
 QY 121 LETVDVFINVKATTTTEKLGPEGRKEGIAVOAVVLIIR 157
 DB 123 LQCDIEGVAVKATTTTEKLGFTGRGEGIAVALLIR 159

RESULT 14

ABP79612
 ID ABP79612 standard; Protein; 174 AA.

XX
 XX
 AC ABP79612;

XX
 XX
 DT 07-MAR-2003 (first entry)

XX
 XX
 DE N. gonorrhoeae amino acid sequence SEQ ID 5754.

XX
 XX
 KW Antibacterial; infection; vaccine; gene therapy.

XX
 XX
 OS Neisseria gonorrhoeae.

XX
 XX
 PN WO200279243-A2.

XX
 XX
 PD 10-OCT-2002.

XX
 XX
 PF 12-FEB-2002; 2002WO-1B02069.

XX
 XX
 PR 12-FEB-2001; 2001GB-0003424.

XX
 XX
 PA (CHIR-) CHIRON SPA.

XX
 XX
 PI Fontana MR, Piza M, Maignani V, Monaci E;

XX
 XX
 DR WPI, 2003-058415/05.

XX
 XX
 DR N-PSDB; AB240582.

XX
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection
 XX
 XX
 PS Disclosure; Page 608; 815pp; English.

CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.

XX
 XX
 SQ Sequence 174 AA;

Query Match 62.9%; Score 507; DB 24; Length 174;
 Best Local Similarity 60.9%; Pred. No. 6e-52;
 Matches 95; Conservative 22; Mismatches 39; Indels 0; Gaps 0;

QY 2 IRVGMGVDRHFRNCDHIIIGVYKIPYKGLFAHSDGDVYTHALADALIGAALGDIGK 61
 DB 18 IRIQGTVDHOUTGRKLLIGVEVPHYTGFIASHDGVNLAHALTDALIGAALGDIGK 77
 QY 62 FPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQAAPKLPVPGKANIADL 121
 DB 78 FPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQAAPKLPVPGKANIADL 137
 QY 122 ETDVDVFINVKATTTTEKLGPEGRKEGIAVOAVVLIIR 157
 DB 138 GIDISCVNIKKTNEKLGVLGRMEGIEQAVALVIR 173

RESULT 15

ABU06061
 ID ABU06061 standard; Protein; 160 AA.

XX
 XX
 AC ABU06061;

XX
 XX
 DT 22-JAN-2003 (first entry)

XX
 XX
 DE N. meningitidis vaccine antigen #40.

XX
 XX
 KW Vaccine; antigen; meningococcal disease; pathogenic bacteria;

XX
 XX
 OS Neisseria meningitidis group B.

XX
 XX
 PN WO200277648-A2.

XX
 XX
 PD 03-OCT-2002.

XX
 XX
 PF 22-MAR-2002; 2002WO-GB01399.

XX
 XX
 PR 22-MAR-2001; 2001GB-0007219.

XX
 XX
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

XX
 XX
 PI (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX
 XX
 PI Robinson AJ, Gorrings AR, Hudson MJ, Bracegirdle P, West DM;

XX
 XX
 PI Oliver KJ, Kroll US, Langford PR;

XX
 XX
 DR WPI, 2003-018958/01.

XX
 XX
 DR N-PSDB; ABX09898.

XX
 XX
 PT Identifying an antigen for manufacturing a vaccine against
 PT meningococcal infection, comprises contacting antibodies with
 PT polypeptides, detecting polypeptide-antibody complexes, and identifying
 PT bound polypeptides as antigens -

XX
 XX
 PS Claim 37; Page 237-238; 310pp; English.

